

UNITED STATES

**Title: PREPARATION OF THIOREDOXIN AND  
THIOREDOXIN REDUCTASE ON OIL BODIES**

**Inventors: Maurice M. Moloney and Bipin K. Dalmia**

**Title: PREPARATION OF THIOREDOXIN AND THIOREDOXIN REDUCTASE PROTEINS ON OIL BODIES**

**5 CROSS-REFERENCE TO RELATED APPLICATIONS**

The present application is a continuation-in-part of USSN 09/210,843 that was filed December 15, 1998, which is a continuation-in-part of USSN 08/846,021 that was filed April 25, 1997 (now U.S. Patent No. 5,948,682), which is a continuation-in-part of USSN 08/366,783 that was filed on December 10, 1994 (now U.S. Patent No. 5,650,554), which is a continuation-in-part of USSN 08/142,418 that was filed November 16, 1993 (now abandoned), which is a continuation-in-part of USSN 07/659,835 that was filed on February 22, 1991 (now abandoned), all of which are incorporated by reference in their entirety.

**FILED OF THE INVENTION**

15 The present invention relates to the production of proteins such as thioredoxin and thioredoxin reductase on oil bodies.

**BACKGROUND OF THE INVENTION**

Many very diverse methods have been tested for the production of recombinant molecules of interest and commercial value. Different 20 organisms that have been considered as hosts for foreign protein expression include single celled organisms such as bacteria and yeasts, cells and cell cultures of animals, fungi and plants and whole organisms such as plants, insects and transgenic animals.

The use of fermentation techniques for large-scale production of 25 bacteria, yeasts and higher organism cell cultures is well established. The capital costs associated with establishment of the facility and the costs of maintenance are negative economic factors. Although the expression levels of proteins that can be achieved are high, energy inputs and protein purification costs can greatly increase the cost of recombinant protein production.

30 The production of a variety of proteins of therapeutic interest has been described in transgenic animals, however the cost of establishing substantial manufacturing is prohibitive for all but high value proteins.

Numerous foreign proteins have been expressed in whole plants and selected plant organs. Methods of stably inserting recombinant DNA into plants have become routine and the number of species that are now accessible to these methods has increased greatly.

5 Plants represent a highly effective and economical means to produce recombinant proteins as they can be grown on a large scale with modest cost inputs and most commercially important species can now be transformed. Although the expression of foreign proteins has been clearly demonstrated, the development of systems with commercially viable levels of  
10 expression coupled with cost effective separation techniques has been limited.

The production of recombinant proteins and peptides in plants has been investigated using a variety of approaches including transcriptional fusions using a strong constitutive plant promoter (e.g., from cauliflower mosaic virus (Sijmons et al., 1990, Bio/Technology, 8:217-221); transcriptional fusions with  
15 organ specific promoter sequences (Radke et al., 1988, Theoret. Appl. Genet., 75:685-694); and translational fusions which require subsequent cleavage of a recombinant protein (Vanderkerckove et al., 1989, Bio/Technology, 7:929-932).

Foreign proteins that have been successfully expressed in plant cells include proteins from bacteria (Fraley et al., 1983, Proc. Natl. Acad. Sci. USA, 20 80:4803-4807), animals (Misra and Gedamu, 1989, Theor. Appl. Genet., 78:161-168), fungi and other plant species (Fraley et al., 1983, Proc. Natl. Acad. Sci. USA, 80:4803-4807). Some proteins, predominantly markers of DNA integration, have been expressed in specific cells and tissues including seeds (Sen Gupta-Gopalan et al., 1985, Proc. Natl. Acad. Sci. USA, 82:3320-3324); Radke et al., 1988,  
25 Theor. Appl. Genet., 75:685-694). Seed specific research has been focused on the use of seed-storage protein promoters as a means of deriving seed-specific expression. Using such a system, Vanderkerckove et al., (1989, Bio/Technol., 7:929-932) expressed the peptide leu-enkephalin in seeds of *Arabidopsis thaliana* and *Brassica napus*. The level of expression of this peptide was quite low and it  
30 appeared that expression of this peptide was limited to endosperm tissue.

It has been generally shown that the construction of chimeric genes which contain the promoter from a given regulated gene and a coding

sequence of a reporter protein not normally associated with that promoter gives rise to regulated expression of the reporter. The use of promoters from seed-specific genes for the expression of recombinant sequences in seed that are not normally expressed in a seed-specific manner have been described.

5 Sengupta-Gopalan et al., (1985, Proc. Natl. Acad. Sci. USA, 82:3320-3324) reported expression of a major storage protein of french bean, called  $\beta$ -phaseolin, in tobacco plants. The gene expressed correctly in the seeds and only at very low levels elsewhere in the plant. However, the constructs used by Sengupta-Gopalan were not chimeric. The entire  $\beta$ -phaseolin gene including the  
10 native 5'-flanking sequences were used. Subsequent experiments with other species (Radke et al., 1988, Theor. App. Genet. 75:685-694) or other genes (Perez-Grau, L., Goldberg, R.B., 1989, Plant Cell, 1:1095-1109) showed the fidelity of expression in a seed-specific manner in both *Arabidopsis* and *Brassica*. Radke et al., (1988, *vide supra*), used a "tagged" gene i.e., one containing the entire napin  
15 gene plus a non-translated "tag".

The role of the storage proteins is to serve as a reserve of nitrogen during seed germination and growth. Although storage protein genes can be expressed at high levels, they represent a class of protein whose complete three-dimensional structure appears important for proper packaging and storage.  
20 The storage proteins generally assemble into multimeric units which are arranged in specific bodies in endosperm tissue. Perturbation of the structure by the addition of foreign peptide sequences leads to storage proteins unable to be packaged properly in the seed.

In addition to nitrogen, the seed also stores lipids. The storage of  
25 lipids occurs in oil or lipid bodies. Analysis of the contents of lipid bodies has demonstrated that in addition to triglyceride and membrane lipids, there are also several polypeptides/proteins associated with the surface or lumen of the oil body (Bowman-Vance and Huang, 1987, J. Biol. Chem., 262:11275-11279, Murphy et al., 1989, Biochem. J., 258:285-293, Taylor et al., 1990, Planta, 181:18-30 26). Oil-body proteins have been identified in a wide range of taxonomically diverse species (Moreau et al., 1980, Plant Physiol., 65:1176-1180; Qu et al., 1986, Biochem. J., 235:57-65) and have been shown to be uniquely localized in oil-

bodies and not found in organelles of vegetative tissues. In *Brassica napus* (rapeseed, canola) there are at least three polypeptides associated with the oil-bodies of developing seeds (Taylor et al., 1990, *Planta*, 181:18-26).

The oil bodies that are produced in seeds are of a similar size  
5 (Huang A.H.C., 1985, in *Modern Meths. Plant Analysis*, Vol. 1:145-151 Springer-  
Verlag, Berlin). Electron microscopic observations have shown that the oil-  
bodies are surrounded by a membrane and are not freely suspended in the  
cytoplasm. These oil-bodies have been variously named by electron  
microscopists as oleosomes, lipid bodies and spherosomes (Gurr MI., 1980, in  
10 *The Biochemistry of Plants*, 4:205-248, Acad. Press, Orlando, Fla). The oil-bodies  
of the species that have been studied are encapsulated by an unusual "half-unit"  
membrane comprising, not a classical lipid bilayer, but rather a single  
amphophilic layer with hydrophobic groups on the inside and hydrophilic  
groups on the outside (Huang A.H.C., 1985, in *Modern Meths. Plant Analysis*,  
15 Vol. 1:145-151 Springer-Verlag, Berlin).

The numbers and sizes of oil-body associated proteins may vary  
from species to species. In corn, for example, there are two immunologically  
distinct polypeptide classes found in oil-bodies (Bowman-Vance and Huang,  
1988, *J. Biol. Chem.*, 263:1476-1481). Oleosins have been shown to comprise  
20 alternate hydrophilic and hydrophobic regions (Bowman-Vance and Huang,  
1987, *J. Biol. Chem.*, 262:11275-11279). The amino acid sequences of oleosins  
from corn, rapeseed, and carrot have been obtained. See Qu and Huang, 1990, *J.*  
*Biol. Chem.*, 265:2238-2243, Hatzopoulos et al., 1990, *Plant Cell*, 2:457-467,  
respectively. In an oilseed such as rapeseed, oleosin may comprise between 8%  
25 (Taylor et al., 1990, *Planta*, 181:18-26) and 20% (Murphy et al., 1989, *Biochem.J.*,  
258:285-293) of total seed protein. Such a level is comparable to that found for  
many seed storage proteins.

Genomic clones encoding oil-body proteins with their associated  
upstream regions have been reported for several species, including maize (*Zea*  
30 *mays*, Bowman-Vance and Huang, 1987, *J. Biol. Chem.*, 262:11275-11279; and Qu  
Huang, 1990, *J. Biol. Chem.*, 265:2238-2243) and carrot (Hatzopoulos et al., 1990,  
*Plant Cell*, 2:457-467). cDNAs and genomic clones have also been reported for

cultivated oilseeds, *Brassica napus* (Murphy, et al., 1991, *Biochem. Biophys. Acta*, 1088:86-94; and Lee and Huang, 1991, *Plant Physiol* 96:1395-1397), sunflower (Cummins and Murphy, 1992, *Plant Molec. Biol.* 19:873-878) soybean ( Kalinski et al., 1991, *Plant Molec. Biol.* 17: 1095-1098), and cotton (Hughes et al., 1993, 5 *Plant Physiol* 101:697-698). Reports on the expression of these oil-body protein genes in developing seeds have varied. In the case of *Zea mays*, transcription of genes encoding oil-body protein isoforms began quite early in seed development and were easily detected 18 days after pollination. In non- 10 endospermic seeds such as the dicotyledonous plant *Brassica napus* (canola, rapeseed), expression of oil-body protein genes seems to occur later in seed development (Murphy, et al., 1989, *Biochem. J.*, 258:285-293) compared to corn.

A maize oleosin has been expressed in seed oil bodies in *Brassica napus* transformed with a *Zea mays* oleosin gene. The gene was expressed under the control of regulatory elements from a *Brassica* gene encoding napin, a major 15 seed storage protein. The temporal regulation and tissue specificity of expression was reported to be correct for a napin gene promoter/terminator (Lee et al., 1991, *Proc. Natl. Acad. Sci. USA*, 88:6181-6185).

Thus the above demonstrates that oil body proteins (or oleosins) from various plant sources share a number of similarities in both structure and 20 expression. However, at the time of the above references it was generally believed that modifications to oleosins or oil body proteins would likely lead to aberrant targeting and instability of the protein product. (Vande Kerckhove et al., 1989. *Bio/Technology*, 7:929-932; Radke et al., 1988. *Theor. and Applied Genetics*, 75:685-694; and Hoffman et al., 1988. *Plant Mol. Biol.* 11:717-729).

25 Of particular relevance to the present invention are the redox proteins thioredoxin and its reductant thioredoxin reductase. Thioredoxins are relatively small proteins (typically approximately 12 kDa) that belong to the family of thioltransferases which catalyze oxido-reductions via the formation or hydrolysis of disulfide bonds and are widely, if not universally, distributed 30 throughout the animal, plant and bacterial kingdom. In order to reduce the oxidized thioredoxin two cellular reductants provide the reduction equivalents, reduced ferredoxin and NADPH. These reduction equivalents are supplied via

different thioredoxin reductases including the NADPH thioredoxin reductase and ferredoxin thioredoxin reductase. The latter thioredoxin reductase is involved in the reduction of plant thioredoxins designated as TRx and TRm, both of which are involved in the regulation of photosynthetic processes in the 5 chloroplast. The NADPH/thioredoxin active in plant seeds is designated TRh and is capable of the reduction of a wide range of proteins thereby functioning as an important cellular redox buffer.

Thioredoxins have been obtained from several organisms including *Arabidopsis thaliana* (Riveira Madrid et al. (1995) Proc. Natl. Acad. Sci. 10 92: 5620-5624), wheat (Gautier et al. (1998) Eur. J. Biochem. 252: 314-324); *Escherichia coli* (Hoeoeg et al (1984) Biosci. Rep. 4: 917-923) and thermophylic microorganisms such as *Methanococcus jannaschii* and *Archaeoglobus fulgidus* (PCT Patent Application 00/36126). Thioredoxins have also been recombinantly expressed in several host systems including bacteria (Gautier 15 et al. (1998) Eur J. Biochem. 252: 314-324) and plants (PCT Patent Application WO 00/58453) Commercial preparations of *E. coli* sourced thioredoxin are readily available from for example: Sigma Cat No. T 0910 Thioredoxin (*E. coli*, recombinant; expressed in *E. coli*).

NADPH-thioredoxin reductase is a cytosolic homodimeric enzyme 20 comprising typically 300-500 amino acids. Crystal structures of both *E. coli* and plant NADPH-thioredoxin reductase have been obtained (Waksman et al. (1994) J. Mol. Biol. 236: 800-816; Dai et al. (1996) J. Allergy Clin. Immunol. 103: 690-697). NADPH-thioredoxin reductases have been expressed in heterologous hosts, for example the *Arabidopsis* NADPH-thioredoxin reductase has been expressed in 25 *E. coli* (Jacquot et al. (1994) J. Mol. Biol. 235: 1357-1363) and wheat (PCT Patent Application 00/58453).

There is a need in the art to further improve the methods for the recombinant expression of thioredoxin and thioredoxin reductase in association with oil bodies.

30 **SUMMARY OF THE INVENTION**

The present invention describes the use of an oil body protein gene to target the expression of a heterologous polypeptide, to an oil body in a

host cell. The unique features of both the oil body protein and the expression patterns are used in this invention to provide a means of synthesizing commercially important proteins on a scale that is difficult if not impossible to achieve using conventional systems of protein production. In a preferred 5 embodiment of the present invention, the heterologous peptide is a thioredoxin or thioredoxin reductase.

In particular, the present invention provides a method for the expression of a thioredoxin or thioredoxin reductase by a host cell said method comprising:

- 10    a) introducing into a host cell a chimeric nucleic acid sequence comprising:
  - 1) a first nucleic acid sequence capable of regulating the transcription in said host cell of
  - 2) a second nucleic acid sequence, wherein said second sequence encodes a fusion polypeptide and comprises (i) a nucleic acid sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the fusion polypeptide to a lipid phase linked in reading frame to (ii) a nucleic acid sequence encoding thioredoxin or thioredoxin reductase; and
  - 3) a third nucleic acid sequence encoding a termination region functional in the host cell; and
- 20    b) growing said host cell to produce the fusion polypeptide.

In a preferred embodiment the oil body protein is an oleosin. Preferred host cells are plant cells, bacterial cells and yeast cells.

The present invention also provides a chimeric nucleic acid 25 sequence encoding a fusion polypeptide, capable of being expressed in association with an oil body of a host cell comprising:

- 1) a first nucleic acid sequence capable of regulating the transcription in said host cell of
- 2) a second nucleic acid sequence, wherein said second sequence encodes a fusion polypeptide and comprises (i) a nucleic acid sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the fusion polypeptide to a lipid phase

linked in reading frame to (ii) a nucleic acid sequence encoding a thioredoxin or thioredoxin reductase; and

- 3) a third nucleic acid sequence encoding a termination region functional in the host cell.

5 In a preferred embodiment the oil body protein is an oleosin. Preferred host cells are plant cells, bacterial cells and yeast cells.

The present invention also includes a fusion polypeptides encoded by a chimeric nucleic acid sequence comprising (i) a nucleic acid sequence encoding a sufficient portion of an oil body protein to provide targeting of the 10 fusion polypeptide to an oil body linked in reading frame to (ii) a nucleic acid sequence encoding a thioredoxin or thioredoxin reductase.

The invention further provides methods for the separation of a thioredoxin or thioredoxin reductase from host cell components by partitioning of the oil body fraction and subsequent release of the thioredoxin or thioredoxin 15 reductase via specific cleavage of the thioredoxin or thioredoxin reductase - oil body protein fusion. Optionally a cleavage site may be located prior to the N- terminus and after the C-terminus of the thioredoxin or thioredoxin reductase protein allowing the fusion polypeptide to be cleaved and separated by phase separation into its component peptides. This production system finds utility in 20 the production of many proteins and peptides such as those with pharmaceutical, enzymic, rheological and adhesive properties.

The methods described above are not limited to thioredoxin or thioredoxin reductase produced in plant seeds as oil body proteins may also be found in association with oil bodies in other cells and tissues. Additionally the 25 methods are not limited to the recovery of thioredoxin or thioredoxin reductase produced in plants because the extraction and release methods can be adapted to accommodate oil body protein-thioredoxin/thioredoxin reductase protein fusions produced in any cell type or organism. An extract containing the fusion protein is mixed with additional oleosins and appropriate tri-glycerides and 30 physical conditions are manipulated to reconstitute the oil-bodies. The reconstituted oil-bodies are separated by floatation and the recombinant

thioredoxin or thioredoxin reductase released by the cleavage of the junction with the oil body protein.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a schematic representation of the types of oil body protein fusions that are contemplated as methods of the invention for the fusion of oil-body protein genes with genes encoding foreign polypeptides. IA is a C-terminal fusion of a desired polypeptide to a oil body protein; IB is an N-terminal fusion of a desired polypeptide to oil body protein; IC is an internal fusion of a desired polypeptide within oil body protein; and ID is an inter-dimer translational fusion of desired polypeptide enclosed between two substantially complete oil body protein targeting sequences. Each fusion is shown in a linear diagrammatic form and in the configuration predicted when specifically associated with the oil body. In both the linear and oil body associated form, the oil body coding sequence that specifically targets the protein to the oil body is shown as a single thin line, a solid circle represents a protease recognition motif; a corkscrew line represents a native C- or N-terminal of a oil body protein and a inserted coding region is represented by an open box. The oil body is represented as a simple circle.

Figure 2 shows the nucleotide sequence (SEQ.ID.NO.1) and deduced amino acid sequence (SEQ.ID.NO.2 and NO.3) of an oil-body protein gene that codes for a 18 KDa oleosin from *Arabidopsis thaliana*. The intron sequence is printed in lower case. The predicted amino acid sequence is shown in single letter code.

Figure 3 shows a schematic representation of the construction of pOleoP1.

Figure 4 shows the nucleotide sequence (SEQ.ID.NO.4) of a *B. napus* oleosin cDNA clone and the predicted amino acid sequence (SEQ.ID.NO.5).

Figure 5 describes the construction of a oleosin/GUS fusion for expression in *E.coli*.

Figure 6 shows the nucleotide sequence (SEQ.ID.NO.6) and deduced amino acid (SEQ.ID.NOS.7 and 8) sequence of the 2.7 kbp *HindIII*

fragment of pSBSOTPNT containing the oleosin-chymosin fusion gene. Indicated in bold (nt 1625-1631) is the *NcoI* site containing the methionine start codon of the prochymosin sequence. The preceding spacer sequence (nt 1608-1630), replacing the oleosin stopcodon is underlined.

5       Figure 7 shows a schematic drawing of plasmid M1830. The plasmid was constructed by replacing the *Ura3* gene from pVT102-U (Gene 52: 225-233, 1987) with the *Leu2* gene.

10      Figure 8 shows a schematic drawing of plasmid M1830oleoGUS. A *BamHI*-GUS-*hindIII* fragment was inserted into the multiple cloning site of M1830, resulting in M1830GUS. A *B. napus* oleosin cDNA was furnished with *BamHI* sites at the 5' and 3' ends of the gene and inserted in frame and in the right orientation in the *BamHI* site of M1830GUS yielding plasmid M1830oleoGUS.

15      Figure 9 shows a comparison of the published NADPH thioredoxin reductase sequence (SEQ.ID.NO.36) (ATTHIREDB Jacquot et al. J. Mol. Biol. (1994) 235 (4):1357-63.) with the sequence isolated in this report (SEQ.ID.NO.37).

20      Figure 10 shows a nucleotide sequence (SEQ.ID.NO.37) and deduced amino acid sequence (SEQ.ID.NO.38) of the NADPH thioredoxin reductase sequence isolated in this report.

25      Figure 11 shows a comparison of the deduced amino acid sequence of the published NADPH thioredoxin reductase sequence (SEQ.ID.NO.39) (ATTHIREDB Jacquot et al. J. Mol. Biol. (1994) 235 (4):1357-63) with the sequence isolated in this report (TR) (SEQ.ID.NO.38).

30      Figure 12 shows the nucleotide sequence of the phaseolin promoter-Arabidopsis Trxh-phaseolin terminator sequence (SEQ.ID.NO.40). The Trxh coding sequence and its deduced amino acid sequence is indicated (SEQ.ID.NO.41). The phaseolin promoter corresponds to nucleotide 6-1554, and the phaseolin terminator corresponds to nucleotide sequence 1905-3124. The promoter was furnished with a *PstI* site (nt 1-6) and the terminator was furnished with a *HindIII* site (nt 1898-1903) and a *KpnI* site (nt 3124-3129) to facilitate cloning.

Figure 13 shows the nucleotide sequence (SEQ.ID.NO.42) of the phaseolin promoter-oleosin Trxh-phaseolin terminator sequence. The oleosin-Trxh coding sequence and its deduced amino acid sequence (SEQ.ID.NOS.:43 and 44) is indicated. As in Figure 12, the phaseolin promoter corresponds to 5 nucleotide 6-1554. The sequence encoding oleosin corresponds to nt 1555-2313, the intron in this sequence (nt 1908-2147) is indicated in italics. The Trxh coding sequence corresponds to nt 2314-2658. The phaseolin terminator corresponds to nucleotide sequence 2664-3884. As in Figure 12 the synthetic PstI, HindIII and KpnI sites are also indicated.

10 Figure 14 shows the nucleotide sequence of the phaseolin promoter - Trxh oleosin-phaseolin terminator sequence (SEQ.ID.NO.45). The Trxh oleosin- coding sequence and its deduced amino acid sequence are indicated (SEQ.ID.NOS.46 and 47). As in Figures 12 and 13, the phaseolin promoter corresponds to nucleotide 6-1554. The Trxh coding sequence 15 corresponds to nt 1555-1896. The sequence encoding oleosin corresponds to nt 1897-2658, the intron in this sequence (nt 2250-2489) is indicated in italics. The phaseolin terminator corresponds to nucleotide sequence 2664-3884. As in Figures 12 and 13 the synthetic PstI, HindIII and KpnI sites are also indicated.

20 Figure 15 shows the nucleotide sequence of the phaseolin promoter - thioredoxin reductase - phaseolin terminator sequence (SEQ.ID.NO.48). The thioredoxin reductase coding sequence and its deduced amino acid sequence is indicated (SEQ.ID.NO.49). The phaseolin promoter corresponds to nucleotide 6-1554. The thioredoxin reductase coding sequence 25 corresponds to nt 1555-2556. The phaseolin terminator corresponds to nucleotide sequence 2563-3782. The synthetic PstI, HindIII and KpnI sites are also indicated.

30 Figure 16 shows the nucleotide sequence of the phaseolin promoter- oleosin thioredoxin reductase-phaseolin terminator sequence (SEQ.ID.NO.50). The oleosin- thioredoxin reductase coding sequence and its deduced amino acid sequence is indicated (SEQ.ID.NOS.51 and 52). The phaseolin promoter corresponds to nucleotide 6-1554. The sequence

encoding oleosin corresponds to nt 1555-2313, the intron in this sequence (nt 1980-2147) is indicated in italics. The thioredoxin reductase coding sequence corresponds to nt 2314-3315. The phaseolin terminator corresponds to nucleotide sequence 3321-4540. The synthetic PstI, HindIII and KpnI sites are 5 also indicated.

Figure 17 shows the nucleotide sequence of the phaseolin promoter - thioredoxin reductase oleosin - phaseolin terminator sequence (SEQ.ID.NO.53). The thioredoxin reductase coding sequence and its deduced amino acid sequence is indicated (SEQ.ID.NOS.54 and 55). The phaseolin 10 promoter corresponds to nucleotide 6-1554. The thioredoxin reductase coding sequence corresponds to nt 1555-2553. The sequence encoding oleosin corresponds to nt 2554-3315, the intron in this sequence (nt 2751-3146) is indicated in italics. The phaseolin terminator corresponds to nucleotide sequence 3321-4540. The synthetic PstI, HindIII and KpnI sites are also 15 indicated.

Figures 18A and B are a gel and Western blot, respectively, showing the analysis of total seed extracts (Lane 1 and 2) and oil body protein extract (Lane 3) of wt Arabidopsis (Lane 1) and Arabidopsis transformed with pSBS2510 (oleosin-thioredoxin). Panel A; coomassie stained gel, Panel B; 20 Western blot treated with a monoclonal antibody raised against Arabidopsis oleosin followed by an alkaline phosphatase linked secondary antibody and NBT/BCIP color reaction. Indicated are the most abundant native oleosin (red arrow) and the oleosin-thioredoxin fusion protein (green arrow).

Figures 19A and B are a gel and Western blot, respectively, 25 showing the analysis of total seed extracts (Lane 2) and oilbody protein extract (Lane 1 and 3) of wt Arabidopsis (Lane 1) and Arabidopsis transformed with pSBS2521 (thioredoxin-oleosin, lane 2 and 3). Panel A; coomassie stained gel, Panel B; Western blot treated with a polyclonal antibody raised against Arabidopsis thioredoxin protein followed by an 30 alkaline phosphatase linked secondary antibody and NBT/BCIP color reaction. Panel C; Western blot treated with a monoclonal antibody raised against Arabidopsis oleosin followed by an alkaline phosphatase linked

secondary antibody and NBT/BCIP color reaction. Indicated is the most abundant native oleosin (red arrow) and the thoredoxin oleosin fusion protein (blue arrow).

Figures 20A and B are a gel and Western blot, respectively,  
5 showing the analysis of total seed extracts of wt Arabidopsis (lane 1) and Arabidopsis transformed with pSBS2520 ("free" thioredoxin, lane 2, 3, 4, 5, 6, 7, 8) Panel A; coomassie stained gel, Panel B; Western blot treated with a polyclonal antibody raised against Arabidopsis thioredoxin protein followed  
10 by an alkaline phosphatase linked secondary antibody and NBT/BCIP color reaction. Indicated is the band reacting with the anti-thioredoxin antibody. A strong signal can be detected in lane 3, 5, 6, 7, 8 and no signal can be detected in the wt extract (Lane 1) and two seed extracts derived from 2 different transgenic lines (lane 2, 4). The lack of detectable thioredoxin expression (as shown in lane 2, 4) could be due to either a position effect or we are looking  
15 or these seeds could be derived from a non-trangenic (escape or false-positive) plant.

Figures 21A and B are Western blots showing the analysis of total seed of wt Arabidopsis (wt) and Arabidopsis transformed with pSBS2529 (thioredoxin reductase-oleosin, lane 1, 2 and 3). Panel A; Western blot treated  
20 with a polyclonal antibody raised against Arabidopsis thioredoxin reductase protein followed by an alkaline phosphatase linked secondary antibody and NBT/BCIP color reaction. Panel B; Western blot treated with a monoclonal antibody raised against Arabidopsis oleosin followed by an alkaline phosphatase linked secondary antibody and NBT/BCIP color reaction.  
25 Indicated is the most abundant native oleosin and the thoredoxin reductase oleosin fusion protein (blue arrow).

Figure 22 is a Western blot showing the analysis of total seed extracts of wt Arabidopsis (wt) and Arabidopsis transformed with pSBS2527 ("free" thioredoxin reductase, Lane 1, 2, 3). Indicated is a Western blot treated  
30 with a polyclonal antibody raised against Arabidopsis thioredoxin reductase protein followed by an alkaline phosphatase linked secondary antibody and

NBT/BCIP color reaction. Indicated is the overexpressed thioredoxin reductase in Lane 1, 2 and 3.

Figures 23A and B are a gel and Western blot, respectively, showing the analysis of total seed extracts (Lane 1 and 2) and oilbody protein 5 extract (Lane 3) of wt Arabidopsis (Lane 1) and Arabidopsis transformed with pSBS2531 (oleosin-Thioreoxin reductasePanel A; Coomassie stained gel, Panel B; Western blot treated with a monoclonal antibody raised against Arabidopsis oleosin followed by an alkaline phosphatase-linked secondary antibody and NBT/BCIP color reaction. Indicated are the most abundant 10 native oleosin (red arrow) and the oleosin-DMSR fusion protein (green arrow). The oilbodies as shown in lane 3 were not washed. As a result some (contaminating) seed proteins can be seen in the oilbody extract as well. However the most abundant proteins in this extract are native oleosin and 15 oleosin-DMSR fusion protein. As expected the wt seed extract (lane 1) showed reactivity only with the native oleosin.

#### DETAILED DESCRIPTION OF THE INVENTION

In accordance with the subject invention, methods and compositions are provided for a novel means of production of heterologous proteins and peptides that can be easily separated from host cell components. In 20 accordance with further embodiments of the invention methods and compositions are provided for novel uses of recombinant proteins produced by said methods.

In accordance with one aspect of the subject invention, methods and compositions are provided for a novel means of production of heterologous 25 proteins and peptides in host cells that are easily separated from other host cell components. Purification of the protein, if required, is greatly simplified. The nucleic acid encoding the heterologous peptide may be part or all of a naturally occurring gene from any source, it may be a synthetic nucleic acid sequence or it may be a combination of naturally occurring and synthetic sequences. The 30 subject method includes the steps of preparing an expression cassette comprising a first nucleic acid sequence capable of regulating the transcription of a second nucleic acid sequence encoding a sufficient portion of an oil body

protein gene to provide targeting to an oil body and fused to this second nucleic acid sequence a third nucleic acid sequence encoding the polypeptide of interest; delivery and incorporation of the expression cassette into a host cell; production of a transformed organism or cell population in which the chimeric gene 5 product is expressed and recovery of a chimeric gene protein product through specific association with an oil body. The heterologous peptide is generally a foreign polypeptide normally not expressed in the host cell or found in association with the oil-body.

In particular, the present invention provides a method for the 10 expression of a heterologous polypeptide by a host cell said method comprising:

- a) introducing into a host cell a chimeric nucleic acid sequence comprising:
  - 1) a first nucleic acid sequence capable of regulating the transcription in said host cell of
  - 2) a second nucleic acid sequence, wherein said second sequence encodes a fusion polypeptide and comprises (i) a nucleic acid sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the fusion polypeptide to a lipid phase linked in reading frame to (ii) a nucleic acid sequence encoding the heterologous polypeptide; and
  - 3) a third nucleic acid sequence encoding a termination region functional in the host cell; and
- b) growing said host cell to produce the fusion polypeptide.

In a preferred embodiment the oil body protein is an oleosin.

In a further preferred embodiment, the present invention provides a 25 method for the expression of a thioredoxin or thioredoxin reductase by a host cell said method comprising:

- a) introducing into a host cell a chimeric nucleic acid sequence comprising:
  - 1) a first nucleic acid sequence capable of regulating the transcription in said host cell of
  - 2) a second nucleic acid sequence, wherein said second sequence encodes a fusion polypeptide and comprises (i) a nucleic acid sequence encoding a sufficient portion of an oil body protein gene

to provide targeting of the fusion polypeptide to a lipid phase linked in reading frame to (ii) a nucleic acid sequence encoding a thioredoxin or thioredoxin reductase; and

3) a third nucleic acid sequence encoding a termination region functional in the host cell; and

5 b) growing said host cell to produce the fusion polypeptide.

The term "oil body protein" as used herein means a protein that can naturally associate with oil bodies or can be isolated using a standard oil body preparation protocol. Examples of oil body proteins include oleosins and 10 caleosins. An oil body preparation protocol is described in van Rooijen and Moloney, 1995, *Bio/Technology*, 13:72-77. The oil body protein may share sequence homology with other oil body proteins which may be oleosins known 15 in the art.

In a preferred embodiment the oil body protein is a plant oleosin and shares sequence homology with other plant oleosins such as the oleosin 20 isolated from *Arabidopsis thaliana* (Figure 2 and SEQ.ID.NO.2) or *Brassica napus* (Figure 4 and SEQ.ID.NO.5). In another embodiment, the oil body protein is a plant caleosin. Caleosin nucleic acid sequences are also known to the art (Naested et al (2000) *Plant Mol Biol.* 44(4):463-476; Chen et al (1999) *Plant Cell Physiol.* 40(10):1079-1086).

The term "heterologous polypeptide" as used herein means a polypeptide, peptide or protein that is not normally linked or fused to an oil body protein and is not normally expressed in association with oil bodies.

The host cell may be selected from a wide range of host cells 25 including plants, bacteria, yeasts, insects and mammals. In one embodiment the host cell is a plant cell. The use of plants to produce proteins of interest allows exploitation of the ability of plants to capture energy and limited nutrient input to make proteins. The scale and yield of material afforded by production in plants allows adaptation of the technology for use in the production of a variety 30 of polypeptides of commercial interest. The plant may be selected from various plant families including *Brassicaceae*, *Compositae*, *Euphorbiaceae*, *Leguminosae*, *Linaceae*, *Malvaceae*, *Umbelliferae* and *Graminae*.

In another embodiment the host cell is a bacterial cell. Bacterial host cells suitable for carrying out the present invention include *E. coli*, *B. subtilis*, *Salmonella typhimurium* and *Staphylococcus*, as well as many other bacterial species well known to one of ordinary skill in the art. Representative examples 5 of bacterial host cells include JM109 ATCC No. 53323 and DH5 (Stratagene, LaJolla, California). Suitable bacterial expression vectors preferably comprise a promoter which functions in the host cell, one or more selectable phenotypic markers, and a bacterial origin of replication. Representative promoters include the LacZ, the  $\beta$ -lactamase (penicillinase) and lactose promoter system (see 10 Chang et al., *Nature* 275:615, 1978), the trp promoter (Nichols and Yanofsky, *Meth in Enzymology* 101:155, 1983) and the tac promoter (Russell et al., *Gene* 20:231, 1982).

In another embodiment, the host cell is a yeast cell. Yeast and fungi host cells suitable for carrying out the present invention include, among 15 others *Saccharomyces cerevisiae*, the genera *Pichia* or *Kluyveromyces* and various species of the genus *Aspergillus*. Suitable expression vectors for yeast and fungi include, among others, YC<sub>p</sub>50 (ATCC No. 37419) for yeast, and the amdS cloning vector pV3 (Turnbull, *Bio/Technology* 7:169, 1989). Protocols for the transformation of yeast are also well known to those of ordinary skill in the art. 20 For example, transformation may be readily accomplished either by preparation of spheroplasts of yeast with DNA (see Hinnen et al., *PNAS USA* 75:1929, 1978) or by treatment with alkaline salts such as LiCl (see Itoh et al., *J. Bacteriology* 153:163, 1983). Transformation of fungi may also be carried out using polyethylene glycol as described by Cullen et al. (*Bio/Technology* 5:369, 1987).

25 The host cell may also be a mammalian cell. Mammalian cells suitable for carrying out the present invention include, among others: COS (e.g., ATCC No. CRL 1650 or 1651), BHK (e.g., ATCC No. CRL 6281), CHO (ATCC No. CCL 61), HeLa (e.g., ATCC No. CCL 2), 293 (ATCC No. 1573) and NS-1 cells. Suitable expression vectors for directing expression in mammalian cells 30 generally include a promoter, as well as other transcriptional and translational control sequences. Suitable promoters include PMSG, pSVL, SV40, pCH 110, MMTV, metallothionein-1, adenovirus Ela, CMV, immediate early,

immunoglobulin heavy chain promoter and enhancer, and RSV-LTR. Protocols for the transfection of mammalian cells are well known to those of ordinary skill in the art. Representative methods include calcium phosphate mediated electroporation, retroviral, and protoplast fusion-mediated transfection (see 5 Sambrook et al., *Molecular Cloning a Laboratory Manual*, 2nd Edition, Cold Spring Harbour Laboratory Press, 1989).

The host cell may also be an insect cell. Insect cells suitable for carrying out the present invention include cells and cell lines from *Bombyx* or *Spodotera* species. Suitable expression vectors for directing expression in insect 10 cells include Baculoviruses such as the *Autographa californica* nuclear polyhedrosis, virus (Miller et al. 1987, in *Genetic Engineering*, Vol. 8 ed. Setler, J.K. et al., Plenum Press, New York) and the *Bombyx mori* nuclear polyhedrosis virus (Maeda et al., 1985, *Nature* 315:592).

The use of an oil body protein as a carrier or targeting means 15 provides a simple mechanism to recover proteins. The chimeric protein associated with the oil body or reconstituted oil body fraction is separated away from the bulk of cellular components in a single step (such as centrifugation size exclusion or floatation); the protein is also protected from degradation during extraction as the separation also reduces contact of the proteins with non-specific 20 proteases.

The invention contemplates the use of heterologous proteins, specifically enzymes, fused to oil body proteins and associated with oil bodies, or reconstituted oil bodies for conversion of substrates in aqueous solutions following mixing of oil body fractions and substrate solutions. Association of the 25 enzyme with the oil body allows subsequent recovery of the enzyme by simple means (centrifugation and floatation) and repeated use thereafter.

In accordance with further embodiments of the invention methods and compositions are provided for the release of heterologous proteins and peptides fused to oleosin proteins specifically associated with isolated oil body or 30 reconstituted oil body fractions. The subject method includes the steps of preparing an expression cassette comprising a first nucleic acid sequence capable of regulating the transcription of a second nucleic acid sequence encoding a

sufficient portion of an oil body protein gene such as oleosin to provide targeting to an oil body and fused to this second nucleic acid sequence via a linker nucleic acid sequence encoding a amino acid sequence cleavable by a specific protease or chemical treatment a third nucleic acid sequence encoding 5 the polypeptide of interest; such that the protein of interest can be cleaved from the isolated oil body fraction by the action of said specific chemical or protease.

For embodiments of the invention wherein the cleavage of heterologous proteins fused to oleosins associated with seed oil bodies is contemplated in germinating seed the expression cassette containing the 10 heterologous protein gene so described above is modified to contain an additional second recombinant nucleic acid molecule comprising a first nucleic acid sequence capable of regulating expression in plants, particularly in germinating seed, more specifically seed embryo or other seed tissue containing oil bodies and under the control of this regulatory sequence a nucleic acid 15 sequence encoding a protease enzyme, specifically a particular protease enzyme capable of cleavage of the fusion protein associated with said oil bodies to release a heterologous protein or peptide from the oil body, and a transcriptional and translational termination region functional in plants. It is desirable that the second recombinant nucleic acid molecule be so constructed 20 such that the first and second recombinant nucleic acid sequences are linked by a multiple cloning site to allow for the convenient substitution of any one of a variety of proteolytic enzymes that may be used to cleave fusion proteins associated with oil bodies.

It is obvious to a person skilled in the art of plant molecular 25 biology, genetics or plant breeding that the equivalent to the above modification to the expression cassette to allow release of proteins and peptides of interest in germinating seeds can be accomplished by other similar means. For example it is possible that the first recombinant nucleic acid molecule and the second recombinant nucleic acid molecule described above may be contained within 30 two independent expression cassettes introduced into the genome of a plant independently. Additionally it is possible to sexually cross a first recombinant plant containing the first recombinant nucleic acid molecule integrated into its

genome with a second recombinant plant with the second recombinant nucleic acid integrated into its genome to produce seed comprising both the first and second nucleic acid molecules.

For embodiments of the invention wherein the heterologous protein is to be produced in and potentially recovered from plant seeds the expression cassette will generally include, in the 5'-3' direction of transcription, a first recombinant nucleic acid sequence comprising a transcriptional and translational regulatory region capable of expression in plants, particularly in developing seed, more specifically seed embryo or other seed tissue that has oil body or triglyceride storage such as pericarp or cuticle, and a second recombinant nucleic acid sequence encoding a fusion peptide or protein comprising a sufficient portion of an oil body specific protein to provide targeting to an oil body, a heterologous protein of interest, and a transcriptional and translational termination region functional in plants. One or more introns may also be present within the oil body specific protein coding sequence or within the coding sequence of the heterologous protein of interest. The fusion peptide or protein may also comprise a peptide sequence linking the oil body specific portion and the peptide or protein of interest that can be specifically cleaved by chemical or enzymatic means. It is desirable that the nucleic acid expression cassette is constructed in such a fashion that the first and second recombinant nucleic acid sequences are linked by a multiple cloning site to allow for the convenient substitution of alternative second recombinant nucleic acid sequences comprising the oil body targeting sequence and any one of a variety of proteins or peptides of interest to be expressed and targeted to oil bodies in seeds.

According to one embodiment of the invention the expression cassette is introduced into a host cell in a form where the expression cassette is stably incorporated into the genome of the host cell. Accordingly it is apparent that one may also introduce the expression cassette as part of a recombinant nucleic acid sequence capable of replication and or expression in the host cell without the need to become integrated into the host chromosome. Examples of this are found in a variety of vectors such as viral or plasmid vectors capable of

replication and expression of proteins in the host cell. One specific example are plasmids that carry an origin of replication that permit high copy number such as the pUC series of *E. coli* plasmids additionally said plasmids modified to contain an inducible promoter such as the *LacZ* promoter inducible by galactose 5 or IPTG.

In an alternative embodiment of the invention nucleic acid is stably incorporated into the genome of the host cell by homologous recombination. Examples of gene targeting by homologous recombination have been described for various cell types including mammalian cells (Mansour et al., 1988, *Nature*, 10 336, 348-352) and plant cells (Miao and Lam, 1995, *Plant Journal*, 7: 359-365). Introduction into the host cell genome of the protein of interest may be accomplished by homologous recombination of the protein of interest in such a fashion that upon recombination an expression cassette is generated which will generally include, in the 5'-3' direction of transcription, a first nucleic acid 15 sequence comprising a transcriptional and translational regulatory region capable of expression in the host cell, a second nucleic acid sequence encoding a fusion protein comprising a sufficient portion of an oil body protein to provide targeting to an oil body and a heterologous protein, and a transcriptional and translational termination region functional in plants.

20 For embodiments of the invention wherein the production and recovery of the heterologous protein is contemplated from non-plant cells the expression cassette so described above is modified to comprise a first recombinant nucleic acid sequence comprising a transcriptional and translational regulatory sequence capable of expression in the intended host production cell 25 or organism. Promoter regions highly active in cells of microorganisms, fungi, insects and animals are well described in the literature of any contemplated host species and may be commercially available or can be obtained by standard methods known to a person skilled in the art. It is apparent that one means to introduce the recombinant molecule to the host cell is through specific infectious 30 entities such as viruses capable of infection of the host modified to contain the recombinant nucleic acid to be expressed.

In a further embodiment of the invention it is contemplated that proteins other than plant oleosins and proteins with homology to plant oleosins that may specifically associate with triglycerides, oils, lipids, fat bodies or any hydrophobic cellular inclusions in the host organism or with reconstituted plant 5 oil bodies may be fused to a recombinant protein and used in the manner contemplated. A system functionally equivalent to plant oleosins and oil bodies has been described in bacteria (Pieper-Fürst et al., 1994, J. Bacteriol. 176:4328 - 4337). Other proteins from additional sources such as, but not limited to; fungi, insects or animals, with equivalent regulatory and targeting properties may be 10 known or discovered by a person skilled in the art.

Of particular interest for transcriptional and translational regulation in plants of the first recombinant nucleic acid molecule is a regulatory sequence (promoter) from an oil body protein gene, preferably an oil body protein gene expressed in dicotyledonous oil seeds. The expression of these 15 genes in dicotyledonous oilseeds was found to occur much earlier than had hitherto been believed as reported in the literature. Thus, the promoters and upstream elements of these genes are valuable for a variety of uses including the modification of metabolism during phases of embryogenesis which precede the accumulation of storage proteins. Alternatively said promoter may also 20 comprise a promoter capable of expression constitutively throughout the plant or a promoter which has enhanced expression within tissues or organs associated with oil synthesis. Of more particular interest is a promoter that expresses an oil body protein to a high level. Many plant species are tetraploid or hexaploid and may contain numerous copies of functional oil body protein 25 genes. As it is preferable to obtain a gene that is controlled by a promoter that expresses at high levels when compared to other oil body protein genes within the same species it may be advantageous to choose a diploid species as a source of oil body protein genes. An example is the diploid cruciferous plant *Arabidopsis thaliana*, wherein only two or three oil body protein genes are 30 detected by southern blot analysis whereas the seeds contain oil body proteins as a high percentage of total protein.

The degree of evolutionary relationship between the plant species chosen for isolation of a promoter and the plant species selected to carry out the invention may not be critical. The universality of most plant genes and promoter function within dicotyledonous species has been amply demonstrated 5 in the literature. Additionally to a certain extent the conservation of function between monocot and dicot genes has also been shown. This is apparent to a person skilled in the art that the function of any given promoter in any chosen species may be tested prior to practising the invention by simple means such as transient expression of marker gene promoter fusions in isolated cells or intact 10 tissues. The promoter region typically comprises minimally from 100 bp 5' to the translational start of the structural gene coding sequence, up to 2.5 kb 5' from the same translational start.

Examples of nucleic acid sequences encoding sequences capable of providing targeting to an oil body protein are oleosins genes obtainable from 15 *Arabidopsis thaliana* or *Brassica napus* which provide for expression of the protein of interest in seed (See Taylor et al., 1990, *Planta* 181:18-26). The necessary regions and amino-acid sequences needed to provide targeting to the oil body reside in the highly hydrophobic central region of oil body proteins. The amino acid sequence necessary to provide targeting to the oil body for *Arabidopsis* 20 *thaliana* oleosins contain amino acids 46-117 shown in SEQ.ID.NO.2. Similarly, the amino acid sequence necessary to provide targeting to the oil body for *Brassica napus* oleosins contains amino acids 60-132 shown in SEQ.ID.NO.5. In a preferred embodiment, the amino acid sequence necessary for targeting 25 additionally contains the N-terminus of the oleosin which includes amino acids 1-45 (SEQ.ID.NO.2) and 1-60 (SEQ.ID.NO.5) for *Arabidopsis* and *Brassica*, respectively.

To identify other oil body protein genes having the desired characteristics, where an oil body protein has been or is isolated, the protein may be partially sequenced, so that a probe may be designed for identifying 30 mRNA. Such a probe is particularly valuable if it is designed to target the coding region of the central hydrophobic domain which is highly conserved among diverse species. In consequence, a nucleic acid or RNA probe for this region

may be particularly useful for identifying coding sequences of oil body proteins from other plant species. To further enhance the concentration of the mRNA, cDNA may be prepared and the cDNA subtracted with mRNA or cDNA from non-oil body producing cells. The residual cDNA may then be used for probing  
5 the genome for complementary sequences, using an appropriate library prepared from plant cells. Sequences which hybridize to the cDNA under stringent conditions may then be isolated.

In some instances, as described above, the use of an oil body protein gene probe (conserved region), may be employed directly for screening  
10 a genomic library and identifying sequences which hybridize to the probe. The isolation may also be performed by a standard immunological screening technique of a seed-specific cDNA expression library. Antibodies may be obtained readily for oil-body proteins using the purification procedure and antibody preparation protocol described by Taylor et al. (1990, *Planta*, 181:18-  
15 26). cDNA expression library screening using antibodies is performed essentially using the techniques of Huynh et al. (1985, in *DNA Cloning*, Vol. 1, a Practical Approach, ed. D.M. Glover, IRL Press, pp. 49-78). Confirmation of sequence is facilitated by the highly conserved central hydrophobic region (see Figure 1). DNA sequencing by the method of Sanger et al. (1977, *Proc. Natl.  
20 Acad. Sci. USA*, 74:5463-5467) or Maxam and Gilbert (1980, *Meth. Enzymol.*, 65:497-560) may be performed on all putative clones and searches for homology performed. Homology of sequences encoding the central hydrophobic domain is typically 70%, both at the amino-acid and nucleotide level between diverse species. If an antibody is available, confirmation of sequence identity may also  
25 be performed by hybrid-select and translation experiments from seed mRNA preparations as described by Sambrook et al. (1990, *Molecular Cloning*, 2nd Ed., Cold Spring Harbour Press, pp. 8-49 to 8-51).

cDNA clones made from seed can be screened using cDNA probes made from the conserved coding regions of any available oil body protein gene  
30 (e.g., Bowman-Vance and Huang, 1987, *J. Biol. Chem.*, 262:11275-11279). Clones are selected which have more intense hybridization with seed nucleic acids as compared to seedling cDNAs. The screening is repeated to identify a particular

cDNA associated with oil bodies of developing seeds using direct antibody screening or hybrid-select and translation. The mRNA complementary to the specific cDNA is absent in other tissues which are tested. The cDNA is then used for screening a genomic library and a fragment selected which hybridizes to the 5 subject cDNA. Of particular interest for transcriptional and translational regulation in plants of said second recombinant nucleic acid molecule is a regulatory sequence (promoter) from a gene expressed during the germination of seeds and the early stages of growth of a seedling, specifically a gene showing high levels of expression during the stage of mobilization of stored seed 10 reserves, more specifically the promoter sequence from the glyoxisomal enzymes iso-citrate lyase or malate synthase. Information concerning genomic clones of iso-citrate lyase and malate synthase from *Brassica napus* and *Arabidopsis* that have been isolated and described has been published (Comai et al., 1989, *Plant Cell* 1: 293-300) and can be used by a person skilled in the art, by 15 the methods described above, to isolate a functional promoter fragment. Other enzymes involved in the metabolism of lipids or other seed reserves during germination may also serve as a source of equivalent regulatory regions.

In order to identify oil body proteins, other than oleosins, oil body preparations such as described in the art for the plants canola (Van Rooijen and 20 Moloney, 1995, *Bio/Technology* 13: 72-77) and peanut (Jacks et al., J.A.O.C.S., 1990, 67: 353-361) and such as described for oil body-like granules in the bacterial species *Rhodococcus ruber* (Pieper-Fürst et al., 1994, *J. Bacteriol.* 176: 4328-4337) may be performed. From such preparations, individual proteins may be readily identified upon electrophoresis on a SDS polyacrylamide gel. Proteins may be 25 extracted from the polyacrylamide gel following the protocol of Weber and Osborn (*J. Biol. Chem.*, 1969, 244: 4406-4412) and polyclonal antibodies against oil body proteins may be obtained using the protocol described by Taylor (1990, *Planta*, 181: 18-26). In order to isolate the corresponding cDNA clone, a cDNA expression library may then be screened with the antibody using techniques 30 familiar to a skilled artisan (see for example: Huynh et al., 1985, in *DNA cloning*, Vol. 1, a Practical Approach, ed. D.M. Glover, IRL Press, pp 49-78).

For production of recombinant protein oleosin fusions in heterologous systems such as animal, insect or microbial species, promoters would be chosen for maximal expression in said cells, tissues or organs to be used for recombinant protein production. The invention is contemplated for use 5 in a variety of organisms which can be genetically altered to express foreign proteins including animals, especially those producing milk such as cattle and goats, invertebrates such as insects, specifically insects that can be reared on a large scale, more specifically those insects which can be infected by recombinant baculoviruses that have been engineered to express oleosin fusion proteins, 10 fungal cells such as yeasts and bacterial cells. Promoter regions highly active in viruses, microorganisms, fungi, insects and animals are well described in the literature and may be commercially available or can be obtained by standard methods known to a person skilled in the art. It is preferred that all of the transcriptional and translational functional elements of the initiation control 15 region are derived from or obtained from the same gene.

For those applications where expression of the recombinant protein is derived from extrachromosomal elements, one may chose a replicon capable of maintaining a high copy number to maximize expression. Alternatively or in addition to high copy number replicons, one may further 20 modify the recombinant nucleic acid sequence to contain specific transcriptional or translation enhancement sequences to assure maximal expression of the foreign protein in host cells.

The level of transcription should be sufficient to provide an amount of RNA capable of resulting in a modified seed, cell, tissue, organ or 25 organism. The term "modified " is meant a detectably different phenotype of a seed, cell, tissue, organ or organism in comparison to the equivalent non-transformed material, for example one not having the expression cassette in question in its genome. It is noted that the RNA may also be an "antisense RNA" capable of altering a phenotype by inhibition of the expression of a 30 particular gene.

Ligation of the nucleic acid sequence encoding the targeting sequence to the gene encoding the polypeptide of interest may take place in

various ways including terminal fusions, internal fusions, and polymeric fusions. In all cases, the fusions are made to avoid disruption of the correct reading frame of the oil-body protein and to avoid inclusion of any translational stop signals in or near the junctions. The different types of terminal and internal fusions are shown in Figure 1 along with a representation of configurations *in vivo*.

In many of the cases described, the ligation of the gene encoding the peptide preferably would include a linker encoding a protease target motif. This would permit the release of the peptide once extracted as a fusion protein.

10 Potential cleavage sites which could be employed are recognition motifs for thrombin (Leu-Val-Pro-Arg-Gly, SEQ.ID.NO.9) (Fujikawa et al., 1972, Biochemistry 11:4892-4899), of factor Xa (Phe-Glu-Gly-Arg-aa, SEQ. ID NO.10) (Nagai et al., 1985, Proc. Natl Acad. Sci. USA, 82:7252-7255) collagenase (Pro-Leu-Gly-Pro, SEQ.ID.NO.11) (Scholtissek and Grosse, 1988, Gene 62:55-64) or

15 Tobacco Etch Virus (TEV) protease (Glu-Asn-Leu-Tyr-Phe-Gln-Gly SEQ.ID.NO.12) (Dougherty et al., 1989, Virology, 172: 302). Additionally, for uses where the fusion protein contains a peptide hormone that is released upon ingestion, the protease recognition motifs may be chosen to reflect the specificity of gut proteases to simplify the release of the peptide.

20 For those uses where chemical cleavage of the polypeptide from the oil body protein fusion is to be employed, one may alter the amino acid sequence of the oil body protein to include or eliminate potential chemical cleavage sites. For example, one may eliminate the internal methionine residues in the *Arabidopsis* oleosin at positions 11 and 117 by site directed mutagenesis to construct a gene that encodes a oleosin that lacks internal methionine residues.

25 By making a N-terminal fusion with the modified oleosin via the N-terminal methionine residue already present in the *Arabidopsis* oleosin, one may cleave the polypeptide of interest by the use of cyanogen bromide providing there are no internal methionines in said polypeptide. Similar strategies for other

30 chemical cleavage agents may be employed. It should be noted that a variety of strategies for cleavage may be employed including a combination of chemical modification and enzymatic cleavage.

By appropriate manipulations, such as restriction, chewing back or filling in overhangs to provide blunt ends, ligation of linkers, or the like, complementary ends of the fragments can be provided for joining and ligation. In carrying out the various steps, cloning is employed, so as to amplify the 5 amount of nucleic acid and to allow for analyzing the nucleic acid to ensure that the operations have occurred in proper manner. A wide variety of cloning vectors are available, where the cloning vector includes a replication system functional in *E. coli* and a marker which allows for selection of the transformed cells. Illustrative vectors include pBR322, pUC series, M13mp series, pACYC184, 10 etc for manipulation of the primary nucleic acid constructs. Thus, the sequence may be inserted into the vector at an appropriate restriction site(s), the resulting plasmid used to transform the *E. coli* host, the *E. coli* grown in an appropriate nutrient medium and the cells harvested and lysed and the plasmid recovered. Analysis may involve sequence analysis, restriction analysis, electrophoresis, or 15 the like. After each manipulation the nucleic acid sequence to be used in the final construct may be restricted and joined to the next sequence, where each of the partial constructs may be cloned in the same or different plasmids.

The mode by which the oil body protein and the protein to be expressed are fused can be either a N-terminal, C-terminal or internal fusion. 20 The choice is dependant upon the application. For example, C-terminal fusions can be made as follows: A genomic clone of an oil body protein gene preferably containing at least 100 bp 5' to the translational start is cloned into a plasmid vehicle capable of replication in a suitable bacterial host (e.g., pUC or pBR322 in *E. coli*). A restriction site is located in the region encoding the hydrophilic C- 25 terminal portion of gene. In a plant oil body protein of approximately 18 KDa, such as the *Arabidopsis* oleosin, this region stretches typically from codons 125 to the end of the clone. The ideal restriction site is unique, but this is not absolutely essential. If no convenient restriction site is located in this region, one may be introduced by site-directed mutagenesis. The only major restriction on 30 the introduction of this site is that it must be placed 5' to the translational stop signal of the OBP clone.

With this altered clone in place, a synthetic oligonucleotide adapter may be produced which contains coding sequence for a protease recognition site such as Pro-Leu-Gly-Pro (SEQ.ID.NO.11) or a multimer thereof. This is the 5 recognition site for the protease collagenase. The adaptor would be synthesized in such a way as to provide a 4-base overhang at the 5' end compatible with the restriction site at the 3' end of the oil body protein clone, a 4-base overhang at the 10 3' end of the adaptor to facilitate ligation to the foreign peptide coding sequence and additional bases, if needed, to ensure no frame shifts in the transition between the oil body protein coding sequence, the protease 15 recognition site and the foreign peptide coding sequence. The final ligation product will contain an almost complete oil body protein gene, coding sequence for collagenase recognition motif and the desired polypeptide coding region all 20 in a single reading frame.

A similar approach is used for N-terminal fusions. The hydrophilic 15 N-terminal end of oil-body proteins permits the fusion of peptides to the N-terminal while still assuring that the foreign peptide would be retained on the outer surface of the oil body. This configuration can be constructed from similar starting materials as used for C-terminal fusions, but requires the identification 20 of a convenient restriction site close to the translational start of the oil body protein gene. A convenient site may be created in many plant oil body protein genes without any alteration in coding sequence by the introduction of a single base change just 5' to the start codon (ATG). In plant oil body proteins thus far studied, the second amino acid is alanine whose codon begins with a "G". A-C 25 transition at that particular "G" yields a *Nco I* site. As an illustration of such a modification, the context of the sequences is shown below:

3' . .TC TCA ACA ATG GCA . . . Carrot Oil Body Protein (SEQ.ID.NO.13)  
3' . .CG GCA GCA ATG GCG . . . Maize 18KDa Oil Body Protein  
(SEQ.ID.NO.14)

A single base change at the adenine prior to the 'ATG' would yield 30 in both cases CCATGG which is an *Nco I* site. Thus, modification of this base using the site-directed mutagenesis will introduce a *Nco I* site which can be used directly for the insertion of a nucleic acid coding sequence assuming no other

*Nco I* sites are present in the sequence. Alternatively other restriction sites may be used or introduced to obtain cassette vectors that provide a convenient means to introduce foreign nucleic acid.

The coding sequence for the foreign peptide may require preparation which will allow its ligation directly into the introduced restriction site. For example, introduction of a coding sequence into the *Nco I* site introduced into the oil body protein coding sequences described above may require the generation of compatible ends. This may typically require a single or two-base modification by site-directed mutagenesis to generate an *Nco I* site around the translational start of the foreign peptide. This peptide is then excised from its cloning vehicle using *Nco I* and a second enzyme which cuts close to the translational stop of the target. Again, using the methods described above, a second convenient site can be introduced by site-directed mutagenesis. It has been suggested by Qu and Huang (1990, *supra*) that the N-terminal methionine might be removed during processing of the plant oil body proteins protein *in vivo* and that the alanine immediately downstream of this might be acylated. To account for this possibility, it may be necessary to retain the Met-Ala sequence at the N-terminal end of the protein. This is easily accomplished using a variety of strategies which introduce a convenient restriction site into the coding sequence in or after the Ala codon.

The resultant constructs from these N-terminal fusions would contain an oil body protein promoter sequence, an in-frame fusion in the first few codons of the oil body protein gene of a high value peptide coding sequence with its own ATG as start signal if necessary and the remainder of the oil body protein gene and terminator.

A third type of fusion involves the placing of a high value peptide coding sequence internally to the coding sequence of the oil body protein. This type of fusion requires the same strategy as in N-terminal fusions, but may only be functional with modifications in regions of low conservation, as it is believed that regions of high conservation in these oil body proteins are essential for targeting of the mature protein. A primary difference in this kind of fusion is the necessity for flanking protease recognition sites for the release of the protein.

This means that in place of the single protease recognition site thus far described, it is necessary to have the protein of interest flanked by one or more copies of the protease recognition site.

Various strategies are dependant on the particular use and nucleic acid sequence of the inserted coding region and would be apparent to those skilled in the art. The preferred method would be to use synthetic oligonucleotides as linkers to introduce the high value peptide coding sequence flanked by appropriate restriction sites or linkers. Orientation is checked by the use of an asymmetrically placed restriction site in the high-value peptide coding sequence.

The heterologous polypeptide of interest to be produced as an oleosin fusion by any of the specific methods described herein, may be any peptide or protein. For example, proteins that alter the amino acid content of seeds may be used. These include genes encoding proteins high in essential amino acids or amino acids that are limiting in diets, especially arginine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan and valine. Storage proteins such as the high lysine 10 KDa zein from *Zea mays* or the 2S high methionine Brazil Nut storage protein may be used. Alternatively synthetic or modified storage proteins may be employed such as peptides encoding poly-lysine or poly-phenylalanine or fusions of one or more coding regions high in essential amino acids. Proteins may also encode useful additives for animal feeds. These proteins may be enzymes for modification of phytate content in meal such as phytase, more specifically phytase from novel sources and having novel activities. Proteins may also encode hormones useful for boosting productivity such as growth hormones or bovine somatotropin. Proteins may also encode peptides useful for aquaculture.

Proteins may also be those used for various industrial processes. Examples of such proteins include chitinase, glucose isomerase, collagenase, amylase, xylanase, cellulase, lipase, chymosin, renin or various proteases or protease inhibitors. One may also express proteins of interest to the cosmetic industry such as collagen, keratin or various other proteins for use in formulation of cosmetics. Proteins of use to the food industry may also be

synthesized including sweetener proteins such as thaumatin, and other flavour enhancing proteins. Proteins that have adhesive properties may also be used.

Of particular interest are those proteins or peptides that may have a therapeutic or diagnostic value. These proteins include enzymes, antigens, 5 such as viral coat proteins or microbial cell wall or toxin proteins or various other antigenic peptides, peptides of direct therapeutic value such as interleukin-1- $\beta$ , the anticoagulant hirudin, blood clotting factors and bactericidal peptides, antibodies, specifically a single-chain antibody comprising a translational fusion of the VH or VL chains of an immunoglobulin. Human growth hormone may 10 also be produced. The invention is not limited by the source or the use of the heterologous polypeptide.

Preferred heterologous proteins of the present application are thioredoxin or thioredoxin reductase.

Accordingly, the present invention provides a chimeric nucleic acid 15 sequence, capable of being expressed in association with an oil body of a host cell comprising:

- 1) a first nucleic acid sequence capable of regulating the transcription in said host cell of
- 2) a second DNA sequence, wherein said second sequence encodes a 20 fusion polypeptide and comprises (i) a nucleic sequence encoding a sufficient portion of an oil body protein gene to provide targeting of the fusion polypeptide to a lipid phase linked in reading frame to (ii) a nucleic sequence encoding a thioredoxin or thioredoxin reductase; and
- 3) a third nucleic acid sequence encoding a termination region 25 functional in the host cell.

In the practice of the present invention any protein which is classified as a thioredoxin may be used, including the thioredoxin component of the NADPH thioredoxin system and the thioredoxin present in the ferredoxin/thioredoxin systems also known as TRx and TRm.

30 In the practice of the invention any protein which is capable of reducing thioredoxin may be used, including the NADPH thioredoxin reductase and the ferredoxin-thioredoxin reductase and any other proteins

having the capability of reducing thioredoxin. In preferred embodiments the thioredoxin and thioredoxin reductase are plant derived.

The nucleic acid sequences encoding thioredoxin polypeptides from diverse biological sources including *E. coli* (Hoeoeg et al. (1984) Biosci. Rep.: 4 917-923); *Methanococcus jannaschii* and *Archaeoglobus fulgidus* (PCT Patent Application 00/36126); *Arabidopsis thaliana* (Rivera-Madrid (1995) Proc. Natl. Acad. Sci. 92: 5620-5624); wheat (Gautier et al (1998) Eur. J. Biochem. 252(2): 314-324); tobacco (Marty et al. (1991) Plant Mol. Biol. 17: 143-148); barley (PCT Patent Application 00/58352); rice (Ishiwatari et al. (1995) Planta 195: 456-463); soybean (Shi et al. (1996) Plant Mol. Biol. 32: 653-662); rapeseed (Bower et al. Plant Cell 8: 1641-1650) and calf (Terashima et al. (1999) DNA Seq. 10(3): 203-205) are available and may all be used in accordance with the present invention. Nucleic acid sequences encoding thioredoxin reductase proteins from *Arabidopsis* (Riveira Madrid et al. (1995) Proc. Natl. Acad. Sci. USA 92: 5620-5624), *E. coli* (Russel et al. (1988) J. Biol. Chem. 263: 9015-9019); barley (PCT Patent Application 00/58352 and wheat (Gautier et al., (1998) Eur. J. Biochem. 252: 314-324) are also known and may be used in accordance with the present invention.

Known nucleic acid sequences encoding thioredoxin and thioredoxin reductase proteins may be used to design and construct nucleic acid sequence based probes in order to uncover and identify previously undiscovered nucleic acid sequences encoding thioredoxin or thioredoxin reductase, for example by screening cDNA or genomic libraries. Thus additional nucleic acid sequences may be discovered and used in accordance with the present invention. In embodiments of the invention in which a thioredoxin and a thioredoxin reductase are used the nucleic acid sequence encoding the thioredoxin and thioredoxin reductase may be obtained from separate sources or may be obtained from the same source. In general however it is preferred that the nucleic acid sequence encoding the thioredoxin polypeptide and the nucleic acid sequence encoding the thioredoxin reductase are obtained from the same or a similar biological source. The nucleic acid sequences encoding the thioredoxin or thioredoxin

reductase proteins may be altered to improve expression levels for example by optimizing the nucleic acids sequence in accordance with the preferred codon usage for the particular cell type which is selected for expression of the thioredoxin or thioredoxin reductase, or by altering of motifs known to 5 destabilize mRNAs (see for example: PCT Patent Application 97/02352). Comparison of the codon usage of the thioredoxin or thioredoxin reductase with codon usage of the host will enable the identification of codons that may be changed. For example typically plant evolution has tended towards a 10 preference for CG rich nucleotide sequences while bacterial evolution has resulted in bias towards AT rich nucleotide sequences. By modifying the nucleic acid sequences to incorporate nucleic acid sequences preferred by the host cell expression may be optimized. Construction of synthetic genes by 15 altering codon usage is described in for example PCT patent Application 93/07278. The thioredoxin or thioredoxin reductase may be altered using for example targeted mutagenesis, random mutagenesis (Shiraishi et al. (1998) Arch. Biochem. Biophys. 358: 104-115; Galkin et al. (1997) Protein Eng. 10: 687- 690; Carugo et al. (1997) Proteins 28: 10-28; Hurley et al. (1996) Biochemistry 35: 5670-5678) (and/or by the addition of organic solvent (Holmberg et al. 20 (1999) Protein Eng. 12: 851-856). In embodiments of the invention in which a thioredoxin and thioredoxin reductase are used, the thioredoxin and thioredoxin reductase may be selected by developing a two-dimensional matrix and determining which combination of first and second redox protein 25 is most effective in electron transport using for example a colorometric reduction assay (Johnson et al (1984) J. of Bact. Vol. 158 3:1061-1069, Luthman et al (1982) Biochemistry Vol 21 26:6628-2233). Combinations of thioredoxin and thioredoxin reductase may be tested by determining the reduction of wheat storage proteins and milk storage protein beta-lactoglobulin in vitro (Del Val et al. (1999) J. Allerg. Clin. Immunol. 103: 690-697).

The termination region which is employed will be primarily one of 30 convenience, since in many cases termination regions appear to be relatively interchangeable. The termination region may be native to the transcriptional initiation region, may be native to the nucleic acid sequence encoding the

polypeptide of interest, or may be derived from another source. Convenient termination regions for plant cell expression are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. Termination signals for expression in other organisms are well known 5 in the literature.

A variety of techniques are available for the introduction of nucleic acid into host cells. For example, the chimeric nucleic acid constructs may be introduced into host cells obtained from dicotyledonous plants, such as tobacco, and oleaginous species, such as *Brassica napus* using standard *Agrobacterium* 10 vectors by a transformation protocol such as that described by Moloney et al., 1989, *Plant Cell Rep.*, 8:238-242 or Hinchee et al., 1988, *Bio/Technol.*, 6:915-922; or other techniques known to those skilled in the art. For example, the use of T-DNA for transformation of plant cells has received extensive study and is amply 15 described in EPA Serial No. 120,516; Hoekema et al., 1985, Chapter V, In: *The Binary Plant Vector System* Offset-drukkerij Kanters B.V., Albllasserdam; Knauf, et al., 1983, *Genetic Analysis of Host Range Expression by Agrobacterium*, p. 245, In: *Molecular Genetics of the Bacteria-Plant Interaction*, Puhler, A. ed., Springer-Verlag, NY; and An et al., 1985, *EMBO J.*, 4:277-284. Conveniently, explants may 20 be cultivated with *A. tumefaciens* or *A. rhizogenes* to allow for transfer of the transcription construct to the plant cells. Following transformation using *Agrobacterium* the plant cells are dispersed in an appropriate medium for selection, subsequently callus, shoots and eventually plantlets are recovered. The *Agrobacterium* host will harbour a plasmid comprising the *vir* genes necessary 25 for transfer of the T-DNA to the plant cells. For injection and electroporation, (see below) disarmed Ti-plasmids (lacking the tumour genes, particularly the T-DNA region) may be introduced into the plant cell.

The use of non-*Agrobacterium* techniques permits the use of the constructs described herein to obtain transformation and expression in a wide variety of monocotyledonous and dicotyledonous plants and other organisms. 30 These techniques are especially useful for species that are intractable in an *Agrobacterium* transformation system. Other techniques for gene transfer include biolistics (Sanford, 1988, *Trends in Biotech.*, 6:299-302), electroporation (Fromm

et al., 1985, Proc. Natl. Acad. Sci. USA, 82:5824-5828; Riggs and Bates, 1986, Proc. Natl. Acad. Sci. USA 83 5602-5606 or PEG-mediated DNA uptake (Potrykus et al., 1985, Mol. Gen. Genet., 199:169-177).

In a specific application, such as to *Brassica napus*, the host cells  
5 targeted to receive recombinant nucleic acid constructs typically will be derived  
from cotyledonary petioles as described by Moloney et al., 1989, Plant Cell Rep.,  
8:238-242). Other examples using commercial oil seeds include cotyledon  
transformation in soybean explants (Hinchee et al., 1988, Bio/technology, 6:915-  
922) and stem transformation of cotton (Umbeck et al., 1981, Bio/technology,  
10 5:263-266).

Following transformation, the cells, for example as leaf discs, are  
grown in selective medium. Once shoots begin to emerge, they are excised and  
placed onto rooting medium. After sufficient roots have formed, the plants are  
transferred to soil. Putative transformed plants are then tested for presence of a  
15 marker. Southern blotting is performed on genomic nucleic acid using an  
appropriate probe, for example an *A. thaliana* oleosin gene, to show that  
integration of the desired sequences into the host cell genome has occurred.

The expression cassette will normally be joined to a marker for  
selection in plant cells. Conveniently, the marker may be resistance to a  
20 herbicide, eg phosphinotricin or glyphosate, or more particularly an antibiotic,  
such as kanamycin, G418, bleomycin, hygromycin, chloramphenicol, or the like.  
The particular marker employed will be one which will allow for selection of  
transformed cells compared with cells lacking the introduced recombinant  
nucleic acid.

25 The fusion peptide in the expression cassette constructed as  
described above, expresses at least preferentially in developing seeds.  
Accordingly, transformed plants grown in accordance with conventional ways,  
are allowed to set seed. See, for example, McCormick et al. (1986, Plant Cell  
Reports, 5:81-84). Northern blotting can be carried out using an appropriate  
30 gene probe with RNA isolated from tissue in which transcription is expected to  
occur such as a seed embryo. The size of the transcripts can then be compared  
with the predicted size for the fusion protein transcript.

Oil-body proteins are then isolated from the seed and analyses performed to determine that the fusion peptide has been expressed. Analyses can be for example by SDS-PAGE. The fusion peptide can be detected using an antibody to the oleosin portion of the fusion peptide. The size of the fusion 5 peptide obtained can then be compared with predicted size of the fusion protein.

Two or more generations of transgenic plants may be grown and either crossed or selfed to allow identification of plants and strains with desired phenotypic characteristics including production of recombinant proteins. It may be desirable to ensure homozygosity of the plants, strains or lines producing 10 recombinant proteins to assure continued inheritance of the recombinant trait. Methods of selecting homozygous plants are well known to those skilled in the art of plant breeding and include recurrent selfing and selection and anther and microspore culture. Homozygous plants may also be obtained by transformation of haploid cells or tissues followed by regeneration of haploid 15 plantlets subsequently converted to diploid plants by any number of known means, (eg: treatment with colchicine or other microtubule disrupting agents).

The desired protein can be extracted from seed that is preferably homozygous for the introduced trait by a variety of techniques, including use of an aqueous, buffered extraction medium and a means of grinding, breaking, 20 pulverizing or otherwise disrupting the cells of the seeds. The extracted seeds can then be separated (for example, by centrifugation or sedimentation of the brei) into three fractions: a sediment or insoluble pellet, an aqueous supernatant, and a buoyant layer comprising seed storage lipid and oil bodies. These oil bodies contain both native oil body proteins and chimeric oil body proteins, the 25 latter containing the foreign peptide. The oil bodies are separated from the water-soluble proteins and re-suspended in aqueous buffer.

If a linker comprising a protease recognition motif has been included in the expression cassette, a protease specific for the recognition motif is added to the resuspension buffer. This releases the required peptide into the 30 aqueous phase. A second centrifugation step will now re-float the processed oil bodies with their attached proteins and leave an aqueous solution of the released peptide or protein. The foreign protein may also be released from the

oil bodies by incubation of the oil body fraction with a different oil body fraction that contains the specific protease fused to oleosin. In this manner the protease cleavage enzyme is removed with the oil bodies that contained the fusion protein with the protease recognition site leaving a product uncontaminated by 5 protease. The desired peptide may be precipitated, chemically modified or lyophilized according to its properties and desired applications

In certain applications the protein may be capable of undergoing self-release. For example, the proteolytic enzyme chymosin undergoes self-activation from a precursor to an active protease by exposure of the precursor 10 to low pH conditions. Expression of the chymosin precursor/oil body fusion protein to conditions of low pH will activate the chymosin. If a chymosin recognition site is included between the oil body protein and the chymosin protein sequences, the activated chymosin can then cleave the fusion proteins. This is an example of self release that can be controlled by manipulation of the 15 conditions required for enzyme activity. Additional examples may be dependant on the requirement for specific co-factors that can be added when self-cleavage is desired. These may include ions, specific chemical co-factors such as NADH or FADH, ATP or other energy sources, or peptides capable of activation of specific enzymes. In certain applications it may not be necessary to remove the fusion 20 protein from the oil-body protein. Such an application would include cases where the fusion peptide includes an enzyme which is tolerant to N or C-terminal fusions and retains its activity; such enzymes could be used without further cleavage and purification. The enzyme/oil body protein fusion would be contacted with substrate. It is also possible to re-use said oil bodies to process 25 additional substrate as a form of an immobilized enzyme. This specific method finds utility in the batch processing of various substances. The process is also useful for enzymatic detoxification of contaminated water or bodies of water where introduction of freely diffusible enzyme may be undesirable. Said process allows recovery of the enzyme with removal of the oil bodies. It is also possible, 30 if desired, to purify the enzyme - oil body protein fusion protein using an immunoaffinity column comprising an immobilized high titre antibody against the oil body protein.

Other uses for the subject invention are as follows. Oil body proteins comprise a high percentage of total seed protein, thus it is possible to enrich the seed for certain desirable properties such as high-lysine, high methionine, and the like, simply by making the fusion protein rich in the amino-  
5 acid(s) of interest could find utility of particular interest is the modification of grains and cereals which are used either directly or indirectly as food sources for livestock, including cattle, poultry, and humans. It may be possible to include, as the fusion peptide, an enzyme which may assist in subsequent processing of the oil or meal in conventional oilseed crushing and extraction, for example  
10 inclusion of a thermostable lipid-modifying enzyme which would remain active at the elevated crushing temperatures used to process seed and thus add value to the extracted triglyceride or protein product. Other uses of the fusion protein include improvement of the agronomic health of the crop. For example, an insecticidal protein or a portion of an immunoglobulin specific for an agronomic  
15 pest such as a fungal cell wall or membrane, could be coupled to the oil body protein thus reducing attack of the seed by a particular plant pest.

It is possible that the polypeptide/protein will itself be valuable and could be extracted and, if desired, further purified. Alternatively the polypeptide/protein or even the mRNA itself may be used to confer a new  
20 biochemical phenotype upon the developing seed. New phenotypes could include such modifications as altered seed-protein or seed oil composition, enhanced production of pre-existing desirable products or properties and the reduction or even suppression of an undesirable gene product using antisense, ribozyme or co-suppression technologies (Izant and Weintraub, 1984, Cell 36:  
25 1007-1015, Hazelhoff and Gerlach, 1988, Nature 334:585-591, Napoli, et al., 1990, Plant Cell, 2:279-289). While one embodiment of the invention contemplates the use of the regulatory sequence in cruciferous plants, it is possible to use the promoter in a wide variety of plant species given the wide conservation of oleosin genes. For example, the promoter could be used in various other  
30 dicotyledonous species as well as monocotyledonous plant. A number of studies have shown the spatial and temporal regulation of dicot genes can be conserved when expressed in a monocotyledonous host. The tomato *rbcS* gene (Kyozuka

et al, 1993, Plant Physiol. 102:991-1000) and the Pin2 gene of potato (Xu et al, 1993 Plant Physiol. 101:683-687) have been shown to function in a monocotyledonous host consistent with their expression pattern observed in the host from which they were derived. Studies have also indicated expression from 5 some dicotyledonous promoters in monocotyledonous hosts can be enhanced by inclusion of an intron derived from a monocotyledonous gene in the coding region of the introduced gene (Xu et al, 1994, Plant Physiol. 106:459-467). Alternatively, given the wide conservation of oleosin genes, it is possible for the skilled artisan to readily isolate oleosin genes from a variety of host plants 10 according to the methodology described within this specification.

It is expected that the desired proteins would be expressed in all embryonic tissue, although different cellular expression can be detected in different tissues of the embryonic axis and cotyledons. This invention has a variety of uses which include improving the intrinsic value of plant seeds by 15 their accumulation of altered polypeptides or novel recombinant peptides or by the incorporation or elimination of a metabolic step. In its simplest embodiment, use of this invention may result in improved protein quality (for example, increased concentrations of essential or rare amino acids), improved lipid quality by a modification of fatty acid composition, or improved or elevated 20 carbohydrate composition. The invention may also be used to control a seed phenotype such as seed coat color or even the development of seed. In some instances it may be advantageous to express a gene that arrests seed development at a particular stage, leading to the production of "seedless" fruit or seeds which contain large amounts of precursors of mature seed products. 25 Extraction of these precursors may be simplified in this case.

Other uses include the inclusion of fusion proteins that contain antigens or vaccines against disease. This application may be particularly relevant to improvements in health care of fish or other wildlife that is not readily assessable by conventional means as the crushed seed can be converted 30 directly into a convenient food source. Other uses include the addition of phytase to improve the nutritional properties of seed for monogastric animals through the release of phosphate from stored phytate, the addition of

chlorophyllase to reduce undesirable chlorophyll contamination of seed oils, especially canola oil and addition of enzymes to reduce anti-metabolites, pigments or toxins from seeds. Additionally the fusion protein may comprise, an insecticidal or fungicidal protein such as magainin or secropin or a portion of an 5 immunoglobulin specific for an agronomic pest, such as a fungal cell wall or membrane, coupled to the oil body protein thus improving seed resistance to pre and post harvest spoilage.

Applications for the use of chimeric proteins associated with the oil body fraction include as above enzymes that are tolerant of N or C-terminal 10 fusions and retain activity. Enzymes associated with oil body suspensions can be mixed with simple or complex solutions containing enzyme substrates. After conversion of substrates to products the enzyme oleosin fusion is readily recovered by centrifugation and floatation and can be reused an indefinite 15 number of times.

15

#### EXAMPLES

The following examples are offered by way of illustration and not by limitation.

**Example 1: Isolation of Plant Oleosin Gene.** Oil body proteins can be isolated from a variety of sources. The isolation of a oil body protein gene (oleosin) from 20 the plant species *Arabidopsis thaliana* is described herein. Similar methods may be used by a person skilled in the art to isolate oil body proteins from other sources. In this example, a *Brassica napus* oleosin gene (described by Murphy et al, 1991, *Biochim Biophys Acta* 1088:86-94) was used to screen a genomic library of *A. thaliana* (cv. Columbia) constructed in the Lamda cloning vector EMBL3A 25 (Obtained from Stratagene Laboratories) using standard techniques. The screening resulted in the isolation of a EMBL 3A clone (referred to as clone 12.1) containing a 15 kb genomic fragment which contains a oleosin gene from *A. thaliana*. The oleosin gene coding region is contained within a 6.6 kb *Kpn I* restriction fragment of this 15 kb fragment. The 6.6kb *KpnI* restriction fragment 30 was further mapped and a 1.8 kb *Nco I* / *Kpn I* fragment containing the oleosin gene including approximately 850 nucleotides of 5' sequence, the complete coding sequence and the 3' region was isolated. This 1.8 kb fragment was end

filled and subcloned in the *Sma I* site of RFM13mp19. The 1.8 kb insert was further digested with a number of standard restriction enzymes and subcloned in M13mp19 for sequencing. Standard cloning procedures were carried out according to Sambrook et al. (Molecular Cloning: A Laboratory Manual 2nd ed., 5 1989, Cold Spring Harbour Laboratory Press.) The nucleotide sequence was determined and the 1.8 kb sequence of the *A. thaliana* oleosin gene is presented in Figure2 and SEQ.ID.NO.1. This particular DNA sequence codes for a 18 KDa *A. thaliana* oleosin gene. The coding region contains a single intron. This gene was used for the construction of recombinant protein expression vectors. The 10 gene may also be used for screening of genomic libraries of other species.

**Example 2: Modification of a Native Oleosin for Expression of Heterologous Proteins.** The DNA fragment described in example 1 that contains the oleosin gene and regulatory elements was incorporated into an expression cassette for use with a variety of foreign/alternative genes. The following illustrates the 15 modification made to the native *A. thaliana* oleosin gene, especially the promoter and coding region, in order to use this gene to illustrate the invention. It is contemplated that a variety of techniques can be used to obtain recombinant molecules, accordingly this example is offered by way of illustration and not limitation. The *A. thaliana* oleosin gene described in example 1 was cloned as a 20 1803 bp fragment flanked by *Nco I* and *Kpn I* sites in a vector called pPAW4. The plasmid pPAW4 is a cloning vehicle derived from the plasmid pPAW1 which is a Bluescript plasmid (Clonetech Laboratories) containing a *Brassica napus* Acetolactate synthase (ALS) gene (Wiersma et al., 1989, Mol Gen Genet. 219:413-420). To construct pPAW4, the plasmid pPAW1 was digested with *Kpn I*. The 25 digested DNA was subjected to agarose gel electrophoresis and the fragment that contained the Bluescript plasmid vector backbone and a 677 base pair portion of the *B. napus* ALS gene was isolated and religated. This plasmid contains the following unique restriction sites within the insert: *Pst I*, *Nco I*, *Hind III* and *Kpn I*. This plasmid was called pPAW4. The 1803 bp *Nco I* - *Kpn I* 30 *Arabidopsis* oleosin gene fragment was cloned between the *Nco I* and *Kpn I* sites in pPAW4. The resultant plasmid contained in addition to the Bluescript plasmid sequences, a 142 bp *Pst I* - *Nco I* fragment derived from the *B. napus* ALS gene

and the entire 1803 bp *Arabidopsis* oleosin gene. The 142 bp *Pst I* - *Nco I* fragment is present only as a "stuffer" fragment as a result of the cloning approach and is not used in oleosin expression constructs.

The resultant plasmid was used to further modify the *Arabidopsis* oleosin gene. Site-directed mutagenesis was used to introduce nucleotide changes at positions -2, -1 and +4 in the DNA sequence shown in figure 2. The changes made were: A to T (nucleotide position -2); A to C (nucleotide position -1) and G to A (nucleotide position +4). These nucleotide changes create a 6 nucleotide *Bsp H1* restriction endonuclease site at nucleotide positions -2 to +4.

5 The *Bsp H1* site (T/CATGA) encompasses the ATG initiation codon and provides a recessed end compatible with *Nco I*. A second modification was made by digestion with the enzymes *Eco RV* and *Msc I* which released a 658 bp fragment containing most of the coding sequence of the native oleosin. This digestion left blunt ends at both the *Eco RV* and *Msc I* sites. The cut vector was recircularized

10 in the presence of an oligonucleotide linker containing the following unique restriction sites: *Hind III*, *Bgl II*, *Sal I*, *Eco RI* and *Cla I*. The recircularized plasmid containing all the 5' regulatory sequences of the oleosin gene, a transcriptional start site and an initiation codon embedded in a *Bsp H1* site. Thirty-one bases downstream of this is a short polylinker containing unique restriction sites. This

15 20 plasmid was called pOleoP1. The restriction map of this construct is shown in figure3.

Introduction of any DNA sequence into pOleoP1, this particular cassette requires that the foreign DNA sequence may have, or be modified to have, a *Bsp H1* or *Nco I* site at the initial ATG position. This will assure

25 conservation of the distance between the "cap" site and the initiator codon. Alternatively restriction site linkers may be added to facilitate insertion into the cassette. The same restriction site can be chosen for the site of insertion of the 3' end of the gene or linkers may be added to introduce appropriate sites. The complete chimeric construct is then excised using the appropriate restriction

30 enzyme(s) and introduced into an appropriate plant transformation vector.

**Example 3: Using the *Arabidopsis* Oleosin Promoter For Controlling Expression in Heterologous Plant Species.** To demonstrate expression of the

oleosin promoter and to determine the amount of 5' regulatory region required for expression in transgenic plants, a small number of DNA constructs were made that contain the 5' transcriptional initiation region of the *Arabidopsis* oleosin gene joined to the coding region for  $\beta$ -glucuronidase (GUS). These 5 constructs were prepared using PCR. The constructs are designated according to the amount of the oleosin 5' region contained, for example, the 2500 construct has approximately 2500 base pairs of the oleosin 5' region. The constructs were introduced into *Brassica napus* and tobacco and the expression of the  $\beta$ -glucuronidase (GUS) gene was measured as described in detail below. The 10 constructs were made using standard molecular biology techniques, including restriction enzyme digestion, ligation and polymerase chain reaction (PCR). As an illustration of the techniques employed, the construction of the 800 construct is described in detail.

In order to obtain a DNA fragment containing approximately 800 15 base pairs from the 5' transcriptional initiation region of the *Arabidopsis* oleosin gene in a configuration suitable for ligation to a GUS coding sequence, PCR was used. To perform the necessary PCR amplification, two oligonucleotide primers were synthesized (Milligen-Bioscience, Cyclone DNA synthesizer). The first primer, the 5' primer, was called GVR10 and had the following sequence (also 20 shown in SEQ.ID.NO.15):

5'-CACTGCAGGAACTCTCTGGTAA-3' (GVR10)

The italicized bases correspond to nucleotide positions -833 to -817 in the sequence reported in Figure 2. The *Pst* 1 site is underlined. The additional nucleotides 5' of this sequence in the primer are not identical to the oleosin gene, 25 but were included in order to place a *Pst* I site at the 5' end of the amplification product.

The second primer, the 3' primer, is designated as ALP 1 and has the following sequence (also shown in SEQ.ID.NO.16):

5'-CTACCCGGGATCCTGTTACTAGAGAGAATG-3' (ALP 1)

30 This primer contains the precise complement (shown in italics) to the sequence reported in Figure 2 from base -13 to -30. In addition, it contains a further 13 bases at the 5' end added to provide two (overlapping) restriction

sites, *Sma* 1 (recognition CCCGGG) and *Bam*H1 (recognition GGATCC), at the 3' end of the amplification product to facilitate cloning of the PCR fragment. Both the *Sma* 1 and *Bam* H1 sites are underlined, the *Bam* H1 site is delineated by a double underline.

5        These two primers were used in a PCR amplification reaction to produce DNA fragment containing the sequence between nucleotides -833 and -13 of the oleosin gene that now contains a *Pst* 1 site at the 5' end and *Sma* 1 and *Bam* H1 sites at the 3' end. The template was the oleosin genomic clone 12.1 described in example 1.

10       The amplification product was called OLEO p800 and was gel purified and digested with *Pst* 1. The digestion product was gel purified and end filled using DNA polymerase Klenow fragment then cut with *Sma* 1 to produce a blunt ended fragment. This fragment was cloned into the *Sma* 1 site of pUC19 to yield the plasmid pUC OLEOp800. This plasmid contained the insert oriented 15 such that the end of the amplified fragment which contained the *Pst* 1 site is proximal to the unique *Hind* III site in the pUC19 cloning vector and the end of the amplified fragment that contains the *Sma* 1 and *Bam* H1 site is proximal to the unique *Eco* RI site in the pUC19. This subclone now contains approximately 800 base pairs of 5' regulatory region from the *Arabidopsis* oleosin gene.

20       The promoter region contained within the plasmid pUC OLEOp800 was fused to the reporter gene GUS. This was accomplished by substituting the oleosin promoter region for a heat shock promoter fused to a GUS gene in the plasmid HspGUS1559. HspGUS1559 is a plasmid used as a binary vector in *Agrobacterium*, derived from the vector pCGN 1559 (MacBride and Summerfeldt, 1990, Plant Molecular Biology, 14, 269-276) with an insert containing heat shock promoter (flanked by *Bam* H1 sites), the  $\beta$ -glucuronidase open reading frame and a nopaline synthase terminator (derived from pB1221, Jefferson RA in Cloning Vectors 1988, Eds. Pouwels P., Enger-Valk BE, Brammer WJ., Elsevier Science Pub BV, Amsterdam section VII, A11). The binary plasmid 25 HspGUS1559 was digested with *Bam*H1 which resulted in the release of the heat shock promoter and permitted the insertion of a *Bam*H1 fragment in its place. 30 HspGUS1559 was then cut with *Bam* H1 to yield a promoter fragment flanked

by *Bam H1* sites. This fragment was cloned into the *Bam H1* sites of the plasmid HspGUS1559 to yield the *Agrobacterium* binary transformation vector pOLEOp800GUS1559. The other constructs were prepared by the same PCR method described above using the appropriate primers for amplifying the -2500 5 fragment, the -1200 fragment, the -600 fragment or the -200 fragment. These plasmids was used to transform *Brassica napus* and tobacco. GUS expression assays (Jefferson R.A., 1987, Plant Mol. Biol. Rep. 5 387-405) were performed on the developing seeds and on non-reproductive plant parts as controls. The 10 results in *Brassica napus* expressed as specific activity of GUS enzyme are shown in Table I. The results in tobacco are shown in Table II. GUS expression reported is an average obtained from approximately five seeds from each of approximately five different transgenic plants.

These results demonstrate that the oleosin fragment from -833 to -13 used in the 800 construct contains sufficient information to direct specific 15 expression of a reporter gene in transgenic *Brassica napus* embryos as early as heart stage and that the *Arabidopsis* oleosin promoter is capable of directing transcription in plants other than *Arabidopsis*.

It should be noted that the specific expression demonstrated here does not depend on interactions with the native terminator of an oleosin gene 3' 20 end. In this example, the 3' oleosin terminator was replaced by a terminator derived from the nopaline synthase gene of *Agrobacterium*. Thus, the sequence in the 800 construct is sufficient to achieve the desired expression profile independent of ancillary sequences.

**Example 4: Use of Oleosin Promoter and Coding Sequences to Direct Fusion 25 Proteins to the Oil Body Fraction of Seeds.** In this example, we have prepared a transgenic plant which expresses, under the control of the oil body promoter, fusion proteins which associate with oilbodies. The enzymatic properties of the inserted coding sequences are preserved while fused to the oleosin. In this example we use the  $\beta$ -glucuronidase enzyme derived from the microorganism *E. coli*. was fused to the oleosin coding region (referred to as a oleosin/GUS 30 fusion) under the control of the *Arabidopsis* oleosin promoter. In order to

create an in-frame GUS fusion with the *Arabidopsis* oleosin, two intermediate plasmids were constructed referred to as pOThromb and pGUSNOS.

The plasmid pOThromb comprises the oleosin 5' regulatory region, the oleosin coding sequence wherein the carboxy terminus of the 5 protein has been modified by addition of a thrombin cleavage site. The plasmid pGUSNOS contains the GUS enzyme coding region followed by the nos terminator polyadenylation signal. These two plasmids were joined to make a fusion protein consisting of the oleosin protein fused to the GUS enzyme by way of a linker peptide that is recognized by the endoprotease thrombin.

10 These plasmids were constructed using PCR and the specific primers shown below. For the construction of pOThromb, a linker oligonucleotide named GVR01 was synthesized having the DNA sequence (shown in SEQ.ID.NO.17) of:

|    |   |    |    |         |
|----|---|----|----|---------|
|    | 10  | 20 | 30 | 40      |
| 15 | 5'AATCCCATGG ATCCTCGTGG AACGAGAGTA GTGTGCTGGC |    |    |         |
|    | CACCACGAGT ACGGTCACGG TC 3'                   |    | 50 | 60      |
|    |   |    |    | (GVR01) |

20 This DNA sequence contains from nucleotides 27-62 sequences complementary to the 3' end of the *Arabidopsis* oleosin coding sequence, from 20 nucleotides 12-26 sequences encoding amino acids that comprise the coding region for a thrombin cleavage site, LVPRGS, and from nucleotides 5-14, the sequence for the restriction sites *Bam HI* and *Nco I*. A second primer referred to as GVR10 was also synthesized and consisting of the following DNA sequence (also shown in SEQ.ID.NO.18):

|    |                               |    |         |
|----|-------------------------------|----|---------|
| 25 | 10                            | 20 |         |
|    | 5'-CACTGCAGGAACCTCTGGTAAGC-3' |    | (GVR10) |

30 This DNA sequence contains from nucleotides 5-24 sequences homologous to the oleosin 5' flanking sequence -834 and -814. These two primers were used to amplify the promoter region (0.8 kb) of the *Arabidopsis* oleosin gene contained in the clone 12.1 described in example 1. The resultant fragment was endfilled and cloned in the *Sma I* site of pUC19. This plasmid was called pOThrom which contained the oleosin promoter region, the oleosin

coding sequence followed by a cleavage site for the enzyme thrombin and restriction sites for the insertion of the  $\beta$ -glucuronidase (hereinafter GUS).

In order to create an in frame GUS fusion with the Arabidopsis oleosin coding region now contained in pOThrom, a GUS gene with the appropriate restriction site was constructed by the use of PCR. An oligonucleotide referred to as GVR20 was synthesized and containing the following DNA sequence (also shown in SEQ.ID.NO.19):

10 20

5'-GAGGATCCATGGTACGTCCCTGTAGAAACC-3' (GVR20)

10 This oligonucleotide contains from nucleotides 9-29, sequences complementary to the GUS gene and from nucleotides 3-12 the sequence for the restriction sites *Bam HI* and *Nco I* to facilitate cloning. In order to create these restriction sites the fourth nucleotide of the GUS sequence was changed from T to G changing the TTA codon (Leu) into GTA (Val). The second primer used  
15 was the universal sequencing primer comprising the DNA sequence (also shown in SEQ.ID.NO.20):

10

5'-GTAAAACGACGGCCAGT-3' (Universal Sequencing Primer)

The GVR20 and the Universal Sequencing Primer were used to  
20 amplify the GUS-nopaline synthase terminator region from the plasmid pBI121  
(Clontech Laboratories). This fragment was endfilled and cloned in the *Sma* I  
site of pUC19. This plasmid was called pGUSNOS.

The plasmid pOThromb was digested with *Pst I* and *Nco I*, pGUSNOS was digested with *Nco I* and *Xba I*. The inserts of both these plasmids were ligated simultaneously into pCGN1559 cut with *Xba I* and *Pst I* to generate plasmid pCGOBPGUS. The plasmid pCGOBPGUS contained in the following order, the *Arabidopsis* oleosin 5' regulatory region, the oleosin coding region, a short amino acid sequence at the carboxy end of the oleosin coding sequence comprising a thrombin protease recognition site, the coding region for the  $\beta$ -glucuronidase gene followed by the nos terminator polyadenylation signal. The fusion protein coded for by this particular DNA construct is designated as an oleosin/GUS fusion protein.

This plasmid pCGOBPGUS was digested with *Pst I* and *Kpn I* cloned into the *PstI* and *Kpn I* sites of pCGN1559 resulting in plasmid pCGOBPGUS which was used as a binary vector in *Agrobacterium* transformation experiments to produce transgenic *B. napus*. Seeds from 5 transgenic *Brassica napus* were obtained and tested for GUS activity. The transformed seeds showed GUS activity specifically associated with the oil body fraction. The results of these experiments are shown in Table III. The data demonstrate specific fractionation of the GUS enzyme to the oil body fraction. This example illustrates the expression and targeting of a bacterial derived 10 enzyme specifically to the oil body fraction of transgenic plants.

One skilled in the art would realize that various modifications can be made to the above method. For example, a constitutive promoter may be used to control the expression of a oleosin/GUS fusion protein. In particular, the 35S promoter may also be used to control the expression of the oleosin/GUS 15 fusion described above by replacing the *Arabidopsis* oleosin promoter with the 35S promoter from CaMV (available from the vector pBI 221.1, Clonetech Laboratories) in the vector pCGOBPGUS. The resultant vector can contain in the following order, the CaMV 35S promoter, the oleosin coding region, a short amino acid sequence at the carboxy end of the oleosin coding sequence 20 comprising a thrombin protease recognition site, the coding region for the  $\beta$ -glucuronidase gene followed by the nos terminator polyadenylation signal. This plasmid can be inserted into Bin 19 and the resultant plasmid may be introduced into *Agrobacterium*. The resulting strain can be used to transform *B.napus*. GUS activity can be measured in the oil body fraction.

25 **Example 5: Cleavage of Oleosin-Fusion Proteins.** In example 4 it was demonstrated that the targeting information contained within the oleosin is sufficient to target the protein oleosin/GUS fusion to the oil body. The oleosin/GUS fusion protein contains an amino acid sequence (LVPRGS SEQ.ID.NO.21), which separates the oleosin from GUS. This sequence is 30 recognized by the protease thrombin, which cleaves this peptide sequence after the arginine (R) amino acid residue. The transgenic seeds containing these oleosin/GUS fusions, were used to demonstrate the general utility of such a

method of cleavage of a foreign peptide from intact oil bodies containing oleosin/foreign peptide-fusions. The oil body fraction that contained the oleosin/GUS fusion was resuspended in thrombin cleavage buffer which consisted of 50 mM Tris (pH 8.0), 150 mM NaCl, 2.5 mM CaCl<sub>2</sub>, 2% Triton X-100 and 0.5 % sarcosyl. Thrombin enzyme was added and the sample was placed for 30 minutes each at 45° C, 50° C and 55° C. Following this incubation oil bodies were recovered and tested for GUS activity. GUS enzymatic activity was found in the aqueous phase following this cleavage and removal of the oil bodies. This is shown in table IV. Western blot analysis confirmed the cleavage 5 of GUS enzyme from the oleosin/GUS fusion protein. This example illustrates the cleavage and recovery of a active enzyme from a oleosin/enzyme fusion 10 following biosynthesis and recovery of the enzyme in the oil body fraction of transgenic seeds.

**Example 6: Use of Fusion Proteins as Reusable Immobilized Enzymes.** In this 15 example, oleosin/GUS fusion proteins that were associated with oilbodies were used as immobilized enzymes for bioconversion of substrates. Advantage was taken of the fact that enzymatic properties are preserved while fused to the oleosin and the oleosin is very specifically and strongly associated with the oil bodies even when the oil bodies are extracted from seeds. In this example it is 20 demonstrated that said fusion enzymes can be used repeatedly and recovered easily by their association with the oil bodies. In order to demonstrate the reusable and stable GUS activity of the transgenic seeds, transgenic oil bodies were isolated from mature dry seeds as follows. The *Brassica napus* transgenic seeds containing a oleosin/GUS fusion protein were ground in extraction buffer 25 A which consists of 0.15 M Tricine-KOH pH 7.5, 10 mM KCl, 1 mM MgCl<sub>2</sub> and 1 mM EDTA, 4°C to which sucrose to a final concentration of 0.6M was added just before use. The ground seeds in extraction buffer were filtered through four layers of cheesecloth before centrifugation for 10 minutes at 5000 x g at 4°C. The oil bodies present as a surface layer were recovered and resuspended in 30 buffer A containing 0.6M sucrose. This solution was overlaid with an equal volume of Buffer A containing 0.1M sucrose and centrifuged at 18,000 x g for 20 minutes. This procedure was repeated twice with the purified oil body fraction

(which contained the oilbodies and oleosin/GUS fusion proteins) and was resuspended in buffer A containing 1mM p-nitrophenyl  $\beta$ -D-glucuronide, a substrate for the GUS enzyme. After incubation, the conversion of the colorless substrate to the yellow p-nitrophenol was used as an indication of GUS activity

5 in the suspensions of transgenic oil bodies. This illustrated the activity of the enzyme is maintained while fused to the oleosin protein and the enzyme is accessible to substrate while attached to the oil bodies. The oil bodies were recovered as described above. No GUS enzyme remained in the aqueous phase after removal of the oil bodies. The oil bodies were then added to fresh

10 substrate. When the oil bodies were allowed to react with fresh substrate, conversion of substrate was demonstrated. This process was repeated four times with no loss of GUS activity. In parallel quantitative experiments, the amount of methyl umbelliferyl glucuronide (MUG) converted to methyl umbelliferone was determined by fluorimetry, and the oil bodies were

15 recovered by flotation centrifugation and added to a new test tube containing MUG. The remaining buffer was tested for residual GUS activity. This procedure was repeated several times. The GUS enzyme showed 100% activity after using four uses and remained stably associated with the oil body fraction. These results are shown in table V. These experiments illustrate the

20 immobilization and recovery of the active enzyme following substrate conversion. The stability of the GUS activity in partially purified oil bodies was established by measuring the GUS activity of the oil body suspension several weeks in a row. The half-life of the GUS activity when the oil-bodies are stored in extraction buffer at 4°C is more than 3 weeks.

25 **EXPRESSION OF OLEOSIN FUSION PROTEINS**

**Example 7: Expression of an Oleosin/IL-1- $\beta$  as a Fusion Protein.** To further illustrate the utility of the invention, the human protein interleukin 1- $\beta$  (IL-1- $\beta$ ) was chosen for biosynthesis according the method. IL-1- $\beta$  consists of 9 amino acids (aa); Val-Gln-Gly-Glu-Glu-Ser-Asn-Asp-Lys (Antoni et al., 1986, J. 30 Immunol. 137:3201-3204 SEQ.ID.NO.22). The strategy for biosynthesis was to place this nine amino acid protein at the carboxy terminus of the native oleosin protein. The strategy further employed the inclusion of a protease recognition

site to permit the cleavage of the IL-1- $\beta$  from the oleosin protein while fused to the oil bodies. In order to accomplish this, a recognition site for the endoprotease Factor Xa was incorporated into the construct. The protease Factor Xa can cleave a protein sequence which contains amino acid sequence ile-5 glu-gly-arg. Cleavage takes place after the arginine residue. Based on these sequences, an oligonucleotide was synthesized which contained 18 nucleotides of the 3' coding region of the *A. thaliana* oleosin (base position 742-759, coding for the last six amino acids of the native protein), an alanine residue (as a result of replacing the TAA stop codon of the native oleosin with a GCT codon for 10 alanine), the coding sequence for the Factor Xa cleavage (four codons for the amino acids ile-glu-gly-arg) followed by the coding sequence for IL-1- $\beta$ . The oligonucleotide further comprised a TAA stop coding after the carboxy terminus lysine residue of IL-1- $\beta$  and adjacent to this stop codon, a *Sal* 1 restriction site was added. The IL-1- $\beta$  coding sequence was designed using optimal codon 15 usage for the *B. napus* and *A. thaliana* oleosin. It is apparent to those skilled in the art that maximal expression is expected when the codon usage of the recombinant protein matches that of other genes expressed in the same plant or plant tissue. This oligonucleotide was inserted into the *Arabidopsis* oleosin gene. The modified oleosin gene was cut with *Pst* 1 and *Sal* 1 and joined to the nos 20 terminator to obtain the plasmid called pCGBPILT. This plasmid contains, in the following order, the *Arabidopsis* oleosin promoter, the oleosin coding sequence, including the intron, and the IL-1- $\beta$  coding region joined at the carboxy terminus of the oleosin protein through a Factor Xa protease recognition site and the nos terminator polyadenylation signal. This construct 25 was inserted into the binary plasmid Bin 19 (Bevan, M., 1984, *Nucl. Acids Res.* 12:8711-8721) and the resultant plasmid was introduced into *Agrobacterium*. The resulting strain was used to transform *B. napus* and tobacco plants.

The *Arabidopsis* oleosin/IL-1- $\beta$  fusion was stably integrated into the genomes of tobacco and *B. napus*. Northern analysis of embryo RNA 30 isolated from different transformed tobacco plants showed the accumulation of *Arabidopsis* oleosin/IL-1- $\beta$  mRNA.

Oil body proteins from transformed tobacco seeds were prepared, and western blotting was performed. An antibody raised against a 22 KDa oleosin of *B. napus*, was used to detect the *Arabidopsis* oleosin/IL-1- $\beta$  fusion in the tobacco seeds. This antibody recognizes all the major oleosins in *B. napus* and *A. thaliana*. In addition, this antibody recognizes the tobacco oleosins. In oleosins extracted from transformed tobacco seeds the antibody recognized a 20 KDa-protein, which represents oleosin/IL-1- $\beta$  fusion oleosin. This fusion protein was not present in the untransformed tobacco seed. These results demonstrate the accumulation of oleosin/IL-1- $\beta$  fusion in tobacco. Similar expression and accumulation is seen in *Brassica napus* transformed with the oleosin/IL-1- $\beta$  fusion gene. These results further exemplify the utility of the method for the expression of heterologous proteins in plants.

**Example 8: Expression of Oleosin/Hirudin Gene Fusion in *B. napus*.** As a further example of the invention, the protein hirudin, derived from the leech (a segmented worm) was synthesized and fused to oleosin. Hirudin is an anti-coagulant which is produced in the salivary glands of the leech *Hirudo medicinalis* (Dodd et al., 1984, FEBS Lett., 65:180-183). The protein is synthesized as a precursor protein (Harvey et al., 1986, Proc. Natl. Acad. Sci. USA 83: 1084-1088) and processed into a 65 amino acid mature protein. The hirudin gene was resynthesized to reflect the codon usage of *Brassica* and *Arabidopsis* oleosin genes and a gene fusion was made with the C-terminal end of the *Arabidopsis* oleosin gene. The gene sequences for oleosin and hirudin were separated by codons for an amino acid sequence encoding a Factor Xa endoprotease cleavage site. The resulting plasmid was called pCGOBHIRT. This plasmid contains, in the following order, the promoter region of the *Arabidopsis* oleosin gene, the coding sequence of the oleosin protein including the intron, a factor Xa cleavage site and the resynthesized hirudin gene followed by the nos terminator polyadenylation signal. This construct was inserted into the binary plasmid Bin 19 and the resultant plasmid was introduced into *Agrobacterium*. The resulting strain was used to transform *B. napus* and tobacco.

The *Arabidopsis* oleosin/hirudin fusion (OBPHIR) was stably integrated into the genomes of *N. tabacum* and *B. napus* respectively. Northern

analysis of embryo RNA isolated from different OBPHIR transformed plants showed the accumulation OBPHIR mRNA in *B.napus* seeds. Monoclonal antibodies raised against hirudin confirmed the stable accumulation of the oleosin/hirudin fusion in the seeds of transformed plants. Transgenic seeds containing an oleosin/hirudin were assayed after a year of storage at room temperature. No degradation of the oleosin/hirudin protein could be observed demonstrating the stability of the hirudin in intact seeds.

The huridin can be cleaved from the oleosin by the use of the Factor Xa cleavage site built into the fusion protein. Upon treatment of the oilbody fraction of transgenic *Brassica napus* seeds, active huridin was released. These results are shown in Table VI. This example illustrates the utility of the invention for the production of heterologous proteins with therapeutic value from non-plant sources.

**Example 9: Fusion of Foreign Proteins to the N-terminus of Oleosin** In this example, a foreign protein was joined to the oleosin coding region via fusion to the N-terminus of the oleosin. As an illustration of the method, the GUS enzyme was fused in-frame to the *Arabidopsis* oleosin coding region described in example 1. In order to accomplish this, four DNA components were ligated to yield a GUS-oleosin fusion under the control of the oleosin promoter. These were: The oleosin 5' regulatory region, the GUS coding region, the oleosin coding region, and the *nos ter* transcription termination region. These four DNA components were constructed as follows:

The first of these components comprised the oleosin promoter isolated by PCR using primers that introduced convenient restriction sites. The 25 5' primer was called OleoPromK and comprised the sequence (also shown as SEQ.ID.NO.23):

5'-CGC GGT ACC ATGG CTA TAC CCA ACC TCG-3'  
                  *Nco1*  
                  *Kpn1*

30

This primer creates a convenient Kpn 1 site in the 5' region of the promoter. The 3' primer comprised the sequence (also shown as SEQ.ID.NO.24):

5'-CGC ATCGATGTTCTTGTACTAGAGAG-3'

Cl<sub>a</sub>1

This primer creates a convenient *Cla* 1 site at the end of the untranslated leader sequence of the oleosin transcribed sequence just prior to the ATG initiation codon in the native oleosin sequence. These two primers were used to amplify a modified promoter region from the native *Arabidopsis* oleosin gene. Following the reaction, the amplification product was digested with *Kpn* 1 and *Cla* 1 to yield a 870bp fragment containing the oleosin promoter and the 5' untranslated leader sequence. This promoter fragment is referred to as Kpn-OleoP-Cla and was ligated in the *Kpn* 2-*Cla* 1 sites of a standard subcloning vector referred to as pBS.

The second DNA component constructed was the GUS coding region modified to introduce the appropriate restriction sites and a Factor Xa cleavage site. In order to accomplish this, the GUS coding region in the vector PBI221 was used as a template in a PCR reaction using the following primers. The 5' primer was called 5'-GUS-Cla which comprised the following sequence (also shown as SEQ.ID.NO.25):

20 5'- GCC ATCGAT<sub>Nde 1</sub> CAT ATG TTA CGT CCT GTA GAA ACC CCA- 3'

The 3' primer was referred to as 3'-GUS-FX-Bam and comprised the following nucleotide sequence (also shown as SEQ.ID.NO.26):

5' CGC **GGATCC** TCT TCC TTC GAT TTG TTT GCC TCC CTG C-3'  
*Bam H1* Factor Xa  
25 encoding DNA sequence  
shown in boldface

This second oligonucleotide also encodes four amino acids specifying the amino acid sequence I-E-G-R, the recognition site for the endoprotease activity of factor Xa. The amplification product of approximately 30 1.8 kb comprises a GUS coding region flanked by a *Cla* 1 site at the 5' end and in place of the GUS termination codon, a short nucleotide sequence encoding the four amino acids that comprise the Factor Xa endoprotease activity cleavage site. Following these amino acid codons is a restriction site for *Bam*H1.

The isolation of the oleosin coding region was also performed using PCR. To isolate this third DNA component, the *Arabidopsis* oleosin genomic clone was used as a template in a reaction that contained the following two primers. The first of these primers is referred to as 5'-Bam-Oleo and has the 5 following sequence (also shown as SEQ.ID.NO.27):

5' CGC GGATCC ATG GCG GAT ACA GCT AGA 3'  
*Bam H1*

The second primer is referred to as 3'-Oleo-Xba and has the 10 following sequence (also shown as SEQ.ID.NO.28):

5' TGC TCT AGA CGA TGA CAT CAG TGG GGT AAC TTA AGT 3'  
*Xba 1*

PCR amplification of the genomic clone yielded an oleosin coding region flanked by a *Bam H1* site at the 5' end and a *Xba 1* site at the 3' end. This 15 coding sequence was subcloned into the *Bam H1* and *Xba 1* site of the subcloning vector pBS.

The fourth DNA component comprised the nopaline synthetase transcriptional termination region (*nos ter*) isolated from the vector pBI 221 as a blunt-ended *Sst 1-EcoRI* fragment cloned into the blunt-ended *Hind III* site of 20 pUC 19. This subclone has a *Xba 1* site at the 5' end and a *Hind III* site at the 3' end.

As a first step to assemble these four DNA components, the oleosin coding region and *nos ter* were first jointed by ligation of the *Bam H1-Xba 1* fragment of the oleosin coding region with the *Xba 1-Hind III* fragment of the 25 *nos ter* into *Bam H1-Hind III* digested pUC 19. This construct yielded a subclone that comprised the oleosin coding region joined to the *nos ter*. As a second step in the assembly of the DNA components, the oleosin promoter region was then joined to the modified GUS coding region by ligation of the *Kpn 1-Cla 1* oleosin promoter fragment to the *Cla 1-Bam H1* fragment of the GUS coding region 30 modified to contain the Factor Xa recognition site and subcloning these ligated fragments into pUC 19 cut with *Kpn 1* and *Bam H1*.

To assemble all four DNA components, the *Kpn 1-Bam H1* oleosin promoter fused to the GUS coding region was ligated with the *Bam H1-Hind III*

oleosin coding region-*nos* *ter* fragment in a tripartite ligation with *Kpn*1-*Hind* III digested *Agrobacterium* binary transformation vector PCGN1559. The resultant transformation vector was called pCGYGON1 and was mobilized into *Agrobacterium tumefaciens* EHA 101 and used to transform *B. napus*.

5 Transformed plants were obtained, transferred to the greenhouses and allowed to set seed. Seeds were analyzed as described by Holbrook et al (1991, Plant Physiology 97:1051-1058) and oil bodies were obtained. Western blotting was used to demonstrate the insertion of the GUS oleosin fusion protein into the oil body membranes. In these experiments, more than 80% of the GUS oleosin

10 fusion protein was associated with the oil body fraction. No degradation of the fusion protein was observed. This example illustrates the utility of the method for the expression and recovery of foreign proteins fused to the N-terminus of oleosin.

**Example 10: Expression of an Oleosin/Chymosin Fusion Protein.** As a further example of the invention, the bovine aspartic protease, chymosin - which is also frequently referred to in the art as rennin - was expressed as an oleosin fusion. Also exemplified here is the cleavage of an oleosin fusion protein by chemical means.

A complementary DNA clone containing a gene of interest may be obtained by any standard technique. For the purpose of this experiment, reverse transcription PCR was used to obtain a full length pre-prochymosin cDNA clone. RNA isolated from calf abomasum was used as the source material for the PCR and primers were designed in accordance with the sequence described by Harris et al. (1982, Nucl. Acids Res., 10: 2177-2187). Subsequently, 25 prochymosin was furnished with an *Nco*I recognition sequence (CCATGG) in such a way that the initiating methionine codon was in frame with the prochymosin cDNA. The Met-prochymosin sequence was ligated in frame to the 3' coding sequence of an *A. thaliana* oleosin genomic sequence oleosin in which the TAA stopcodon had been replaced by a short spacer sequence (encoding 30 LVPRGS SEQ.ID.NO.29) and an *Nco*I site. The complete sequence of a *Hind*III fragment containing the oleosin-spacer-Met-prochymosin sequence is shown in Fig. 6 and SEQ.ID.NO.6. This *Hind*III fragment was joined to a nopaline synthase

terminator and cloned into the binary vector pCGN1559 (McBride and Summerfelt, 1990, *Plant Mol. Biol.* 14: 269-276). The resulting plasmid was called pSBSOPTNT and introduced in *A. tumefaciens*. The resulting bacterial strain was used to transform *B. napus* plants.

5 Oil bodies from transformed *B. napus* plants were prepared and resuspended in 100mM Tris-Cl, pH 8.0. In order to demonstrate chemical cleavage of chymosin from the oleosin-spacer-Met-prochymosin fusion, the pH of the oil body suspension was lowered into two steps to pH 5.5 and pH 3.0, respectively using HCl. Oil bodies were subjected to these acidic conditions for  
10 several hours prior to Western blotting. Western blotting was performed using polyclonal antibodies raised against bovine chymosin and using commercially available chymosin (Sigma) as a positive control. The oleosin-spacer-Met-  
15 prochymosin fusion protein (approximately 62 kDa) could only be detected in oil body protein extracts obtained from transgenic *B. napus* seeds incubated at pH 8.0 and pH 5.0. No mature chymosin (35 kDa) was detected in protein extracts incubated under these conditions. The mature chymosin polypeptide was detected as the predominant molecular species in oil body protein extracts incubated at pH 3.0. In addition, oil body protein extracts incubated at pH 3.0  
20 were the only extracts exhibiting chymosin activity as measured by milk-clotting assay. In protein extracts isolated from untransformed control plants no specific cross-reactivity with anti-chymosin antibodies was detected.

**Example 11: Expression of an Oleosin/Cystatin Fusion Protein.** As a further example of the present invention, the expression of a protein that is toxic to insects is illustrated. The cysteine protease inhibitor, cystatin (OC-I), from *Oryza sativa* was expressed in a germination-specific manner in *Brassica napus* cv. Westar. The strategy for biosynthesis was to place the coding sequence for the complete 11.5 kDa OC-I protein downstream of the isocitrate lysase (ICL) promoter, isolated from *Brassica napus* (Comai et al., 1989, *Plant Cell* 1: 293-300). The ICL promoter has been shown to be functional for several days directly  
25 after germination of the seeds. Thus, this will allow for the pulse release of cystatin only for several days after germination when seedlings are most  
30

susceptible to the feeding of insects such as the flea beetle (*Phyllotreta cruciferae*) or the red turnip beetle (*Entomoscelis americana*).

The 313 bp sequence, encoding OC-I, from the cDNA clone OC 9b (Chen et al., 1992, *Prot. Expr. and Purif.*, 3: 41-49) was amplified by PCR, using 5' and 3' specific primers, designed to introduce *BspHI* and *BamHI* sites for cloning purposes. The resulting fragment was cloned into pITG7, a vector containing the *nos* terminator of transcription. OC-I-nos was amplified from this plasmid by PCR, using the 5' primer specific to the OC-I coding sequence and the Universal primer (Stratagene). The resulting OC-I-nos fragment was cloned into 10 the *SmaI* site of pBS(KS), excised with *BspHI* and *KpnI* and introduced into pUC18-ICL (plasmid containing the ICL promoter) at the *NcoI* and *KpnI* sites. The entire ICL-OC-I-nos cassette was removed by digestion with *PstI*, cloned into the plant binary vector pCGN 1547 (McBride and Summerfelt, 1990, *Plant Mol. Biol.* 14: 269-276) and designated pCGN-ICLOC. This plasmid was 15 introduced into *Agrobacterium tumefaciens* EHA101 and the resulting strain was used to transform *Brassica napus* cv. Westar, using the cut petiole transformation method (Moloney et al., 1989, *Plant Cell Reports* 8: 238-242). Transformation resulted in the stable integration of the ICL-OC-I -nos construct into the genome of *Brassica napus*. Northern blot analysis of poly-A<sup>+</sup> mRNA isolated from 20 seedlings showed the accumulation of OC-I mRNA transcripts between one (1) to four (4) days after germination.

Protein extracts from the cotyledons of transformed *Brassica napus* seedlings were prepared using standard techniques (Sambrook et al., 1989, *Molecular Cloning: a laboratory manual* 2nd ed, Cold Spring Harbor 25 Laboratory Press) and Western blot analysis was performed in order to determine if OC-I protein was produced. A polyclonal antibody raised against the truncated 10 kDa recombinant form of OC-I (Chen et al., 1992) was produced and allowed the detection of the complete OC-I protein (11.5 kDa) in extracts prepared from transformed *Brassica napus* seedlings. The OC-I protein 30 was not detected in ungerminated seeds or in untransformed seeds or seedlings. The expression of OC-I was also found to be tissue specific, with the protein

being found in cotyledons and hypocotyls but absent from roots and the first true leaves.

In order to prove functionality of the OC-I protein produced in the *Brassica napus* seedlings, a proteinase inhibitor assay (Rymerson et al., manuscript in preparation) was performed, using the proteinase papain. OC-I produced in the seedlings was shown to significantly inhibit the activity of papain. The experiments described here, indicate that OC-I protein, cystatin, is produced in a germination and tissue specific manner and acts as a functional proteinase inhibitor in this system.

10 **Example 12: Expression of an Oleosin/Xylanase Fusion Protein.** As a further example of the present invention, the production of an industrial enzyme, xylanase, is illustrated. A variety of industrial applications have been reported for xylanases (Jeffries et al., 1994, TAPPI 77: 173-179; Biely, 1985, Trends Biotechnol. 3: 286-290), including the conversion of the pulp and paper industry 15 waste product xylan to useful monosaccharides.

The *xynC* gene encoding a highly active xylanase from the rumen fungus *Neocallimastix patriciarum* (Selinger et al., 1995, Abstract, 23rd Biennial Conference on Rumen Function, Chicago, Illinois) was joined in-frame to oleosin via a fusion to the C-terminus of the *Arabidopsis* oleosin coding region described in example 1. The *xynC* gene consists of an N-terminal catalytic domain preceded by a signal peptide. The xylanase gene lacking the ATG startcodon and partial signal peptide coding sequence was first amplified by PCR using the following 2 primers (also shown in SEQ.ID.NO.30 and SEQ.ID.NO.31):

25 10 20 30

and

10 20

5'-GGGTGCTCGAGATTCATAATCAATTAT-3'

30 The PCR product was digested with *EcoRI* and *XhoI* and cloned into the *E. coli* expression vector pGEX4T-3 (Pharmacia) and designated pGEXxyn. Following expression and purification of the xylanase-glutathion-S-transferase fusion

protein according to the protocol provided by the manufacturer, polyclonal antibodies against xylanase were obtained from rabbits immunized with thrombin-cleaved, purified recombinant xylanase.

In order to obtain the 1608 bp fragment containing the oleosin promoter and oleosin coding region, the construct pCGYOBPGUSA (van Rooijen and Moloney, 1995, *Plant Physiol.* 109: 1353-1361) was digested with *PstI* and *BamHI*. The xylanase coding region was obtained by digestion of pGEXxyn with *EcoRI* and *XhoI*. The oleosin fragment and xylanase fragment were cloned into pBluescript (pBS), previously digested with *EcoRI* and *XhoI*, resulting in pBSOleXyn. In order to isolate the nopaline synthase (NOS) terminator region containing *XbaI* and *XhoI* cloning sites, the *BamHI* - HindIII fragment from pCGYOBPGUSA containing the NOS terminator sequence was subcloned in pBS to yield the intermediate plasmid pBSNos. Digestion of pBSNos with *XbaI* and *XhoI* and digestion of pBSOleXyn with *PstI* and *XhoI* yielded fragments containing the NOS terminator and the oleosin-xylanase fusion respectively and were ligated into the binary vector pCGN1559 which was digested with *PstI* and *XhoI*. The resulting binary vector containing the recombinant oleosin-xylanase fusion was named pCGOleXyn. Following introduction of pCGOleXyn into *A. tumefaciens*, *B. napus* cv Westar plants were transformed using the method of Moloney et al. (1989, *Plant Cell Rep.* 8: 238-242).

Accumulation of oleosin-XynC fusion protein in oil-bodies of transgenic canola plants was assessed by Western analysis. Probing of total seed protein extracts and oil body protein extracts with anti-XynC antiserum revealed the presence of a predominant band of 70 kDa on Western blots in both extracts. The predicted molecular weight of the oleosin-XynC fusion protein (68.2 kDa) and hence is in good agreement with the observed band. The fusion protein was absent in extracts from untransformed plants.

In order to evaluate functional activity of the oleosin-xylanase fusion proteins, xylanase enzyme assays using remazol brilliant blue-xylan (RBB-xylan) as described by Biely et al. (1988, *Methods Enzymol.* 160: 536-542) were carried out using oilbody immobilized xylanase. Xylanase activity was

found to be associated almost exclusively with the oil body fraction and kinetic parameters were comparable to those of microbially expressed xylanase.

**Example 13: Expression of an Oleosin/Carp Growth Hormone Fusion Protein.**

As a further example of this invention, the production of carp growth hormone (cGH) as an oleosin fusion protein is described. A DNA fragment containing the cGH coding region lacking its 22 amino acid signal sequence was amplified from a plasmid containing on an insert a common carp (*Cyprinus carpio*) growth hormone cDNA (Koren et al., 1989, Gene 67: 309-315) using the PCR in combination with two cGH-specific primers. The amplified cGH fragment was 5 fused in the correct reading frame and 3' to the *A. thaliana* oleosin using pOThromb (van Rooijen, 1993, PhD Thesis, University of Calgary) as a parent plasmid and employing cloning strategies similar to those outlined in the present application in e.g. examples 9 to 11 and well known to a person skilled in the art. In pOThromb a thrombin cleavage site was engineered 3' to the oleosin 10 coding sequence. The oleosin-cGH fusion was introduced into the binary vector pCGN1559 (McBride and Summerfelt, 1990, Plant Mol. Biol. 14: 269-276) and the resulting construct was used to transform *A. tumefaciens*. The *Agrobacterium* 15 strain was employed to transform *B. napus* cv Westar seedlings.

Seeds from transgenic *B. napus* plants were analysed for cGH expression by Western blotting using monoclonal antibodies against cGH. The 20 expected 40 kDa oleosin-cGH fusion protein was specifically detected in oil body protein extracts containing the oleosin-cGH fusion protein. A 22 kDa polypeptide corresponding with cGH could be released from oil bodies upon treatment with thrombin, while no cGH was detected in oil body protein 25 extracts from untransformed control plants.

**Example 14: Expression of an Oleosin/Zein Fusion Protein.** In order to demonstrate the utility of the instant invention for the production of improved meal, a gene specifying high levels of methionine residues, was expressed as an oleosin fusion in *B. napus* seeds. For the purpose of this experiment the gene 30 encoding the corn seed storage protein zein (Kirihsara et al., 1988, Gene 71: 359-370) was used. The zein gene was fused 3' of the oleosin coding sequence and introduced in the binary vector pCGN1559 (McBride and Summerfelt, 1990,

Plant Mol. Biol. 14: 269-276) employing cloning strategies similar to those described in the present application in e.g. examples 9 to 11 and well known to the skilled artisan. The resulting recombinant plasmid was introduced in *A. tumefaciens* and used to transform *B. napus* cotyledonary explants. Amino acid 5 analyses of canola meal of plants transformed with the oleosin-zein fusion construct indicated a significant increase in the levels of methionine in the meal when compared to untransformed plants.

**Example 15: Construction of an Oleosin/Collagenase Protein Vector.** As a further example of the invention, a vector containing an oleosin-collagenase 10 fusion was constructed.

A 2.2 kbp fragment containing the collagenase gene from *vibrio alginolyticus* was PCR amplified from genomic bacterial DNA using primers in accordance with the published sequence (Takeuchi et al., 1992, Biochem. Journal, 281: 703-708). The fragment 2.2 kbp was then subcloned into pUC19 yielding 15 pZAP1. Subsequently, the collagenase gene was introduced into pNOS8 containing the NOS terminator. The collagenase gene was ligated to the oleosin promoter and coding sequence of pThromb (van Rooijen, 1993, PhD Thesis, University of Calgary) containing a thrombin cleavage site and introduced into the binary vector pCGN1559 (McBride and Summerfelt, 1990, Plant Mol. Biol. 14: 20 269-276).

The collagenase construct may be introduced in a transgenic plant containing a second oleosin gene fusion to, for example, a gene encoding the enzyme chitinase isolated from tobacco (Melchers et al., 1994, Plant Journal 5: 469-480) and containing a collagenase recognition sequence engineered between 25 the oleosin sequence and the second fusion protein. Introduction of the two fusion genes may be accomplished by sexual crossing of two lines which each contain one of the fusion genes or by transformation of a plant containing the first construct the second construct.

**EXPRESSION IN PLANT HOSTS**  
30 **Example 16: Expression of Oleosin/GUS Fusions in Various Plant Species.** It is a feature of the present invention that a wide variety of host cells may be employed. In order to illustrate the expression of oleosin fusions in a number of

plant species, the expression of the *A. thaliana* oleosin fused to the reporter gene GUS was assessed in the embryos of nine different plant species, including the monocotelydenous plant species *Zea mays* (corn).

Plasmid pCGYOBPGUS containing the intact *A. thaliana* oleosin gene with a carboxyl terminal fused GUS gene (van Rooijen et al., 1995, Plant Physiol. 109: 1353-1361) was used to transform oilseed embryos of the following plant species: *Brassica napus* (canola), *Helianthus annuus* (sunflower), *Carthamus tinctorius* (safflower), *Glycine max* (soybean), *Ricinus communis* (castor bean), *Linum usitatissimum* (flax), *Gossypium hirsutum* (cotton), *Coriandrum sativum* (coriander) and *Zea mays* (corn). Transformation was accomplished by particle bombardment (Klein et al., 1987, Nature, 327: 70-73) and plasmid pGN, containing a promoterless GUS gene was used as a control. Histochemical GUS staining (Klein et al., 1988, Proc. Natl. Acad. Sci. 85: 8502-8505) of the embryos was used to assess GUS expression.

The embryos of the 9 species transformed with plasmid pCGTYOBPGUS containing the oleosin-GUS fusion gene all exhibited substantial GUS expression as judged by histochemical staining. In contrast, no appreciable levels of GUS activity was detected in embryos transformed with the promoterless GUS construct.

## 20 EXPRESSION IN PROKARYOTES

**Example 17: Isolation of a *B. napus* Oleosin cDNA** The *Arabidopsis* oleosin gene described in Example 1 contains an intron, and as such is not suitable for use in a prokaryotic expression system. In order to express oleosin fusions in a microorganism such as bacteria, a coding sequence devoid of introns must be used. To accomplish this, a *B. napus* cDNA library was made using standard techniques and was used to isolate oleosin cDNAs. Four clones were obtained and were called pcDNA#7, pcDNA#8, pcDNA#10 and pcDNA#12. These cDNA clones were partly sequenced, and one clone pcDNA#8, was sequenced completely. All the clones showed high levels of identity to oleosins. pcDNA#10 was identical to pcDNA#12, but different from pcDNA#8 and pcDNA#7. The deduced amino acid sequence of the insert of pcDNA#8 is very similar to the *Arabidopsis* oleosin and is shown in figure 4. This coding region

of oleosin can be used to isolate other oleosin genes or for expression of oleosin fusions in prokaryotic systems. It also provides a convenient coding region for fusion with various other promoters for heterologous expression of foreign proteins due to the ability of the protein (oleosin) to specifically interact with the  
5 oilbody fraction of plant extracts.

**Example 18: Expression of a Oleosin/GUS Fusion in the Heterologous Host *E. coli*.** In order to further illustrate the invention, an oleosin/GUS gene fusion was expressed in *E. coli* strain JM109. The oleosin cDNA pcDNA#8 described in example 17 was digested with *Nco I* and ligated into the *Nco I* site of pKKGUS, an  
10 expression vector containing the LacZ promoter fused to GUS. The plasmid pKKGUS was constructed by adding the GUS coding region to the vector pKK233 (Pharmacia) to generate the plasmid pKKoleoGUS and the anti-sense construct pKKoeloGUS. This construct is shown in Figure 5. These plasmids were introduced into *E. coli* strain JM109 and expression was induced by IPTG.  
15 The *E. coli* cells were prepared for GUS activity measurements. In bacterial cells containing the vector pKKGUS, strong induction of GUS activity is observed following addition of IPTG. In cells containing pKKoleoGUS similar strong induction of GUS activity was seen following addition of IPTG. In cells containing pKKoeloGUS (GUS in the antisense orientation) no induction over  
20 background was observed following the addition of IPTG. These results suggest that the oleosin/GUS fusion is active in bacteria. Although that activity observed for the fusion product is less than the unfused product, the oleosin coding sequence was not optimized for expression in bacteria. It is apparent to those skilled in the art that simple modification of codons or other sequences  
25 such as ribosome binding sites could be employed to increase expression. The results are summarized in Table VII.

The fusion protein can be isolated from the bulk of the cellular material by utilizing the ability of the oleosin portion of the fusion proteins to specifically associate with oil bodies.

30 **EXPRESSION IN FUNGI**

**Example 19: Expression of an Oleosin/GUS Fusion in the Heterologous Host *Saccharomyces cerevisiae*.** As an example of the utility of the disclosed

invention for expression in fungal systems, an oleosin-GUS fusion was expressed in *S. cerevisiae*. Plasmids pM1830OleoGUS, containing an oleosin-GUS fusion, and control plasmid pM1830 (Figs 7 and 8) were used to transform *S. cerevisiae* strain 1788 (Mata/Mata) an isogenic diploid of EG123 (MAT $\alpha$  *leu2*-3,112 *ura3*-52*trp1**his4**canI*); Kyung and Levin, 1992, Mol. Cel. Biol. 12: 172-182) according to the method of Elbe (1992, Biotechniques: 13: 18-19). Briefly, strain 1788 was grown on YPD (1% yeast extract, 2% peptone, 2% dextrose; Sherman et al., methods in yeast genetics, Cold Spring Harbor Laboratory Press) at 30° C for 1 day. The strain was then transformed with plasmids pM1830 and pM1830OleoGUS. Transformants were selected on synthetic media (SC, Sherman et al. Methods in yeast genetics, Cold Spring Harbor Laboratory Press), lacking leucine at 30° C for 3 days. Individual colonies were grown in SC (minus leucine) for 1 day, reinoculated into fresh medium at equal starting densities ( $OD_{600} = 0.05$ ), then grown to mid-log phase ( $OD_{600} = 2.0-3.0$ ). Cultures were centrifuged at 4,000 rpm for 5 min and the supernatant was removed. The pellet was resuspended in 100 mM Tris-Cl (pH 7.5), 1mM PMSF (phenyl methyl sulphonyl fluoride) and the cells were lysed using a French Press. GUS activity measurements were done according to Jefferson (1987) and protein determination was done as described by Bradford et al. (1976, Anal. Biochem. 72: 248-254).

Western blotting using a polyclonal anti-oleosin antibody revealed the presence of a 90 kDa polypeptide, which is in agreement with the molecular weight deduced from the amino acid sequence of the fusion protein (89.7 kDa). No cross-reactivity was observed in extracts from the untransformed strain or in extract transformed with the control plasmid pM1830. Significant GUS activity could be detected in *S. cerevisiae* cells transformed with pM1830OleoGUS, while no appreciable levels of GUS activity were measured in untransformed cells or cells transformed with pM1830 (table VIII).

#### EXAMPLE 20

#### 30 Isolation of thioredoxin and NADPH thioredoxin reductase genes

An *Arabidopsis* silique cDNA library CD4-12 was obtained from the *Arabidopsis* Biological Resource Centre (ABRC, <http://aims.cps.msu.edu>)

Arabidopsis stock centre and used as a template for the isolation of the thioredoxin h (Trxh) and thioredoxin reductase genes from Arabidopsis. For the isolation of the Trxh gene the following primers were synthesized:

GVR833: 5' **TACCATGGCTTCGGAAGGAAGGA** 3' (SEQ.ID.NO.32)

5 The sequence identical to the 5' end of the Trxh gene as published in Rivera-Madrid et al, (1993) Plant Physiol 102: 327-328, is indicated in bold. Underlined is an NcoI restriction site to facilitate cloning.

GVR834: 5' **GAAAGCTTAAAGCCAAGTGTTTG** 3' (SEQ.ID.NO.33)

10 The sequence complementary to the 3' end of the Trxh gene as published in Rivera-Madrid et al, (1993) Plant Physiol 102: 327-328, is indicated in bold. Underlined is an HindIII restriction site to facilitate cloning.

15 A Polymerase Chain Reaction (PCR) was carried out using GVR833 and GVR834 as primers and the cDNA library CD4-12 as a template. The resulted PCR fragment was isolated, cloned into pBluescript and sequenced. The isolated sequence encoding Trxh was identical to the published Trxh gene sequence (Rivera-Madrid et al, (1993) Plant Physiol 102: 327-328). The pBluescript vector containing the Trxh gene is called pSBS2500

For the isolation of the thioredoxin reductase gene the following primers were synthesized:

20 GVR836: 5' **GGCCAGCACACTACCATGGTCTCGAAACTCAC** 3' (SEQ.ID.NO.34). The sequence identical to the 5' end of the thioredoxin reductase gene as published (ATTHIREDB Jacquot et al, J Mol Biol. (1994) 235 (4):1357-63), is indicated in bold).

GVR837: 5' **TTAAGCTTCAATCACTCTACCTTGCTG** 3' (SEQ.ID.NO.35)

25 A Polymerase Chain Reaction (PCR) was carried out using GVR836 and GVR837 as primers and the cDNA library CD4-12 as a template. The resulted PCR fragment was isolated, cloned into pBluescript and sequenced. The pBluescript vector containing the thioredoxin reductase gene is called pSBS2502.

30 A total of three clones were sequenced, the sequence of each of the three clones were identical to each other. However, as depicted in Figure 9 this sequence indicated several nucleotide differences compared to the

published thioredoxin reductase gene sequence published (ATTHIREDB Jacquot et al, J Mol Biol. (1994) 235 (4):1357-63). The complete coding sequence and its deduced aminoacid sequence is shown in Figure 10. As a result of the nucleotide differences between the published sequence and the sequence 5 isolated in this report, several amino acid changes are also predicted. A comparison of the deduced amino acid sequence of the published NADPH thioredoxin reductase sequence thioredoxin (ATTHIREDB, Jacquot et al, J Mol Biol. (1994) 235 (4):1357-63.) with the sequence isolated in this report (TR) is shown in Figure 11.

10 **EXAMPLE 21**

**Construction of plant expression vectors.**

Expression vectors were constructed to allow for the seed specific over-expression of thioredoxin and NADPH thioredoxin reductase in seeds. Vectors were constructed to allow for over-expression in its natural 15 subcellular location and for accumulation on oilbodies.

*Construction of plant transformation vector pSBS2520.* The Arabidopsis thioredoxin h gene as described in Example 20 was placed under the regulatory control of the phaseolin promoter and the phaseolin terminator derived from the common bean *Phaseolus vulgaris* (Slightom et al 20 (1983) Proc. Natl Acad Sc USA 80: 1897-1901; Sengupta-Gopalan *et al.*, (1985) PNAS USA 82: 3320-3324)). A gene splicing by overlap extension technique (Horton et al (1989) 15: 61--68) was used to fuse the phaseolin promoter to the Trxh gene. Standard molecular biology laboratory techniques (see eg: Sambrook *et al.* (1990) Molecular Cloning, 2<sup>nd</sup> ed. Cold Spring Harbor Press) 25 were used to furnish the phaseolin promoter and terminator with Pst I and HindIII/KpnI sites respectively (see Figure 12). Standard molecular biology laboratory techniques were also used to place the phaseolin terminator downstream from the Trxh gene. The PstI-phaseolin promoter- Trxh-phaseolin terminator-KpnI insert sequence was cloned into the PstI-KpnI sites 30 of pSBS3000 (pSBS3000 is a derivative from the *Agrobacterium* binary plasmid pPZP221 (Hajdukiewicz et al., 1994, Plant Molec. Biol. 25: 989-994). In pSBS3000, the CaMV35S promoter-gentamycin resistance gene-CAMV 35S

terminator of pPZP221 was replaced with parsley ubiquitin promoter-phosphinothricin acetyl transferase gene-parsley ubiquitin termination sequence to confer resistance to the herbicide glufosinate ammonium.) The resulting plasmid is called pSBS2520. The sequence of the phaseolin 5 promoter-Arabidopsis Trxh-phaseolin terminator sequence is shown in Figure 12.

*Construction of plant transformation vector pSBS2510.* The 3' coding sequence of an Arabidopsis oleosin gene (Van Rooijen et al (1992) Plant Mol. Biol.18: 1177-1179). was altered to contain an NcoI site. The NcoI- HindIII 10 fragment from vector pSBS2500 (Example 20) containing the Trxh was ligated to the coding sequence of this Arabidopsis oleosin utilizing this NcoI restriction site . A gene splicing by overlap extension technique (Horton et al (1989) 15: 61--68) was used to fuse the phaseolin promoter (Slightom et al (1983) Proc. Natl Acad Sc USA 80: 1897-1901; Sengupta-Gopalan et al., (1985) 15 PNAS USA 82: 3320-3324) containing a synthetic PstI site (see construction of pSBS2520)) to the coding sequence of the Arabidopsis oleosin. Standard 20 molecular biology laboratory techniques (see eg: Sambrook et al. (1990) Molecular Cloning, 2<sup>nd</sup> ed. Cold Spring Harbor Press) were again used to clone the HindIII KpnI fragment containing the phaseolin terminator (see construction of pSBS2520) downstream of the Trxh gene. The PstI-phaseolin promoter- oleosin- Trxh-phaseolin terminator-KpnI insert sequence was cloned into the PstI-KpnI sites of pSBS3000. The resulting plasmid is called pSBS2510. The sequence of the phaseolin promoter-oleosin Trxh-phaseolin terminator sequence is shown in Figure 13.

*Construction of plant transformation vector pSBS2521.* This vector 25 contains the same genetic elements as the insert of pSBS2510 except the Trxh gene is fused to the 5' end of the oleosin gene. The 3' oleosin coding sequence including its native stopcodon (Van Rooijen et al (1992) Plant Mol. Biol.18: 1177-1179) was furnished with a HindIII cloning site. Again a gene splicing by 30 overlap extension technique (Horton et al (1989) 15: 61--68) was used to fuse the phaseolin promoter to the Trxh gene and to fuse the Trxh gene to the oleosin sequence. Standard molecular biology laboratory techniques (see eg:

Sambrook *et al.* (1990) Molecular Cloning, 2<sup>nd</sup> ed. Cold Spring Harbor Press) were again used to clone the HindIII KpnI fragment containing the phaseolin terminator (see construction of pSBS2520) downstream of the oleosin gene. The PstI-phaseolin promoter-Trxh oleosin- phaseolin terminator-KpnI insert 5 sequence was cloned into the PstI-KpnI sites of pSBS3000. The resulting plasmid is called pSBS2521. The sequence of the phaseolin promoter-Trxh oleosin -phaseolin terminator sequence is shown in Figure 14.

*Construction of plant transformation vector pSBS2527.* The Arabidopsis NADPH thioredoxin reductase gene as described in Example 20 10 was placed under the regulatory control of the phaseolin promoter and the phaseolin terminator derived from the common bean *Phaseolus vulgaris* (Slightom *et al* (1983) Proc. Natl Acad Sc USA 80: 1897-1901; Sengupta-Gopalan *et al.*, (1985) PNAS USA 82: 3320-3324). A gene splicing by overlap 15 extension technique (Horton *et al* (1989) 15: 61-68) was used to fuse the phaseolin promoter to the thioredoxin reductase gene. Standard molecular 20 biology laboratory techniques (see eg: Sambrook *et al.* (1990) Molecular Cloning, 2<sup>nd</sup> ed. Cold Spring Harbor Press) were used to furnish the phaseolin promoter and terminator with PstI and HindIII/KpnI sites 25 respectively (see Figure 12). Standard molecular biology laboratory techniques were also used to place the phaseolin terminator downstream from the thioredoxin reductase gene. The PstI-phaseolin promoter-thioredoxin reductase-phaseolin terminator-KpnI insert sequence was cloned into the PstI-KpnI sites of pSBS3000 The resulting plasmid is called pSBS2527. The sequence of the phaseolin promoter-Arabidopsis thioredoxin reductase- phaseolin terminator sequence is shown in Figure 15.

*Construction of plant transformation vector pSBS2531.* A gene splicing by overlap extension technique (Horton *et al* (1989) 15: 61-68) was used to fuse the phaseolin promoter (Slightom *et al* (1983) Proc. Natl Acad Sc USA 80: 1897-1901; Sengupta-Gopalan *et al.*, (1985) PNAS USA 82: 3320-3324 to 30 the coding sequence of the Arabidopsis oleosin. The same gene splicing technique was used to fuse the oleosin gene to the thioredoxin reductase coding sequence. Standard molecular biology laboratory techniques (see eg:

Sambrook *et al.* (1990) Molecular Cloning, 2<sup>nd</sup> ed. Cold Spring Harbor Press) were again used to clone the HindIII KpnI fragment containing the phaseolin downstream of the thioredoxin reductase gene. The PstI-phaseolin promoter-oleosin-thioredoxin reductase-phaseolin terminator-KpnI insert sequence was 5 cloned into the PstI-KpnI sites of pSBS3000. The resulting plasmid is called pSBS2531. The sequence of the phaseolin promoter-oleosin thioredoxin reductase-phaseolin terminator sequence is shown in Figure 16.

*Construction of plant transformation vector pSBS2529.* This vector contains the same genetic elements as the insert of pSBS2531 except the 10 thioredoxin gene is fused to the 5' end of the oleosin gene. The 3' oleosin coding sequence including its native stopcodon (Van Rooijen *et al.* (1992) Plant Mol. Biol.18: 1177-1179) was furnished with a HindIII cloning site. Again a gene splicing by overlap extension technique (Horton *et al* (1989) 15: 61-68) was used to fuse the phaseolin promoter to the thioredoxin reductase gene 15 and to fuse the thioredoxin reductase gene to the oleosin sequence. Standard molecular biology laboratory techniques (see eg: Sambrook *et al.* (1990) Molecular Cloning, 2<sup>nd</sup> ed. Cold Spring Harbor Press) were again used to clone the HindIII KpnI fragment containing the phaseolin terminator (see construction of pSBS2520) downstream of the oleosin gene. The PstI-phaseolin promoter- thioredoxin reductase oleosin- phaseolin terminator-KpnI insert sequence was cloned into the PstI-KpnI sites of pSBS3000. The resulting 20 plasmid is called pSBS2529. The sequence of the phaseolin promoter-thioredoxin reductase oleosin -phaseolin terminator sequence is shown in Figure 17.

25 Plasmids pSBS2510, pSBS2520, pSBS2521, pSBS2527, pSBS2529 and pSBS2531and were electroporated into *Agrobacterium* strain EHA101. These *Agrobacterium* strains were used to transform *Arabidopsis*. *Arabidopsis* transformation was done essentially as described in "Arabidopsis Protocols; Methods in molecular biology Vol 82. Edited by Martinez-Zapater JM and 30 Salinas J. ISBN 0-89603-391-0 pg 259-266 (1998) except the putative transgenic plants were selected on agarose plates containing 80μM L-phosphinothricine, after they were transplanted to soil and allowed to set seed.

### **EXAMPLE 22**

#### **Polyacrylamide gel electrophoresis and immunoblotting of transgenic seed extracts.**

Source of *Arabidopsis* thioredoxin, thioredoxin reductase and 5 oleosin antibodies. The *Arabidopsis* thioredoxin and thioredoxin reductase genes were cloned in frame in bacterial expression vector pRSETB (Invitrogen) to allow for the overexpression of *Arabidopsis* thioredoxin and thioredoxin reductase proteins. These proteins were purified using standard 10 protocols (see eg Invitrogen protocol) and used to raise antibodies in rabbits using standard biochemical techniques (See eg Current Protocols in Molecular 15 Biology, John Wiley & Sons, N.Y. (1989). The *Arabidopsis* oleosin gene genes was cloned in frame in bacterial expression vector pRSETB (Invitrogen) to allow for the overexpression *Arabidopsis* oleosin protein. This protein was purified using standard protocols (see eg Invitrogen protocol) and used to prepare mouse monoclonal antibodies using standard biochemical techniques 20 (See eg Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989).

*Preparation of total *Arabidopsis* seed extracts for PAGE.* *Arabidopsis* seeds were ground in approximately 20 volumes of 2% SDS, 50 mM Tris-Cl,, 20 this extract was boiled, spun and the supernatant was prepared for polyacrylamide gelectrophoresis (PAGE) using standard protocols.

*Preparation of *Arabidopsis* oil body protein extracts.* *Arabidopsis* seeds were ground in approximately 20 volumes of water and spun in a microfuge. The oilbodies were recovered and washed sequentially with 25 approximately 20 volumes of water, a high stringency wash buffer, containing 8M urea and 100 mM sodiumcarbonate and water. After this last wash the oilbodies are prepared for poly acrylamide gelectrophoresis (PAGE) using standard protocols.

*Analysis of seed and oil body extracts from plants transformed with 30 pSBS2510.* Total seed and oilbody protein extracts from plants transformed with pSBS2510 were loaded onto polyacrylamide gels and either stained with coomassie brilliant blue or electroblotted onto PVDF membranes. The

membranes were challenged with with a polyclonal antibody raised against Arabidopsis thioredoxin, or a monoclonal antibody raised against the Arabidopsis 18.5 kDa oleosin and and visualized using alkaline phosphatase. Expression of the oleosin-thioredoxin results in an additional band of 31.2 5 kDa. The results are shown in Figure 19. The thioredoxin antibodies are immunologically reactive with a band of the right predicted molecular weight (31.2 kDa), the oleosin antibodies are also immunologically reactive with a band of the right predicted molecular weight for the fusion protein (31.2 kDa) in addition to a band corresponding to the native Arabidopsis oleosin (18.5 10 kDa). This indicates that oleosin-thioredoxin is expressed in Arabidopsis seeds and is correctly targeted to oilbodies.

*Analysis of seed and oil body extracts from plants transformed with pSBS2521.* Total seed and oilbody protein extracts from plants transformed with pSBS25121 were loaded onto polyacrylamide gels and either stained with 15 coomassie brilliant blue or electroblotted onto PVDF membranes. The membranes were challenged with with a polyclonal antibody raised against Arabidopsis thioredoxin, or a monoclonal antibody raised against the Arabidopsis 18.5 kDa oleosin and and visualized using alkaline phosphatase. Expression of the thioredoxin-oleosin results in an additional band of 31.2 20 kDa. The results are shown in Figure 19. The thioredoxin antibodies are immunologically reactive with a band of the right predicted molecular weight (31.2 kDa), the oleosin antibodies are also immunologically reactive with a band of the right predicted molecular weight for the fusion protein (31.2 kDa) in addition to a band corresponding to the native Arabidopsis oleosin (18.5 25 kDa). This indicates that thioredoxin-oleosin is expressed in Arabidopsis seeds and is correctly targeted to oilbodies.

*Analysis of seed extracts from plants transformed with pSBS2520.* Total seed extracts from plants transformed with pSBS2520 were loaded onto polyacrylamide gels and either stained with coomassie brilliant blue or 30 electroblotted onto PVDF membranes. The membranes were challenged with with a polyclonal antibody raised against Arabidosis thioredoxin and visualized using alkaline phosphatase. The thioredoxin antibodies are

immunologically reactive with a band of approximately the right predicted molecular weight (12 kDa). Untransformed seeds do not show a detectable thioredoxin band (results not shown). The results of this analysis are shown in Figure 20.

5         *Analysis of seed and oil body extracts from plants transformed with pSBS2529.* Total seed and oilbody protein extracts from plants transformed with pSBS2529 were loaded onto polyacrylamide gels and electroblotted onto PVDF membranes. The membranes were challenged with with a polyclonal antibody raised against *Arabidopsis* thioredoxin reductase, or a monoclonal 10 antibody raised against the *Arabidopsis* 18.5 kDa oleosin and and visualized using alkaline phosphatase. Expression of the thioredoxin reductase -oleosin results in an additional band of 53.8 kDa. The results are shown in Figure 21. The thioredoxin reductase antibodies are immunologically reactive with a band of the right predicted molecular weight for the fusion protein (53.8 15 kDa), the oleosin antibodies are also immunologically reactive with a band of the right predicted molecular weight (53.8 kDa) in addition to a band corresponding to the native *Arabidopsis* oleosin (18.5 kDa). This indicates that thioredoxin reductase-oleosin is expressed in *Arabidopsis* seeds.

20         *Analysis of seed extracts from plants transformed with pSBS2527.* Total seed extracts from plants transformed with pSBS2527 were loaded onto polyacrylamide gels and electroblotted onto PVDF membranes. The membranes were challenged with with a polyclonal antibody raised against *Arabidopsis* thioredoxin reductase and visualized using alkaline phosphatase. The thioredoxin reductase antibodies are immunologically reactive with a 25 band of approximately the right predicted molecular weight for the (35.3 kDa). Untransformed seeds do not show a detectable thioredoxin band. The results of this analysis are shown in Figure 22.

25         *Analysis of seed extracts from plants transformed with pSBS2531.* Figure 23 shows a protein gel and immunoblot of the expression of oleosin-30 DMSR in *Arabidopsis* T2 seeds and correct targeting to *Arabidopsis* oilbodies. The expected molecular weight based on the deduced amino acid sequence is calculated to be 53,817 Da. In the oilbody extract of the putative transgenic

oleosin-thioredoxin reductase sample an extra band of approximately 54 kDa can be seen (Panel A). This band is confirmed to be oleosin- thioredoxin reductase by immunoblotting (Panel B). From the polyacrylamide gel it can be seen that the expression of the oleosin -Thioredoxin reductase is about 5 double compared to the expression of the major 18.5 kDa *Arabidopsis* oleosin. This represents approximately 2-4 % of total seed protein.

#### **ADDITIONAL APPLICATIONS OF THE INVENTION**

The above examples describe various proteins that can be fused to an oil body protein and expressed in oil bodies in plants, bacteria and yeast. The 10 above also provides the methodology to prepare such transgenic plants. Therefore one skilled in the art can readily modify the above in order to prepare fusion proteins containing any desired protein or polypeptide fused to an oil body protein in a variety of host systems. Several examples of other applications of the present invention are provided below.

15 a) **Construction of an Oleosin/Single Chain Antibody Fusion Protein.** As a further example of the invention, an antibody may be expressed in *B. napus*. Prior to the construction of an oleosin gene fusion, the deduced amino acid sequence of the coding region for the antibody may be back-translated using a *B. napus* codon usage table derived from several known *B.* 20 *napus* genes and 'inside-out' recursive PCR (Prodomou & Pearl, 1992, Protein Eng. 5: 827-829) and yielding a synthetic scFv gene.

Gene fusion between the oleosin gene and the antibody gene can be accomplished by joining the synthetic antibody gene to the 5' end of the oleosin gene in a plasmid using cloning strategies well known to a person skilled 25 in the art and similar to those outlined in the subject application in e.g. Examples 9 to 11. The insert from the plasmid may be cloned into the binary vector pCGN1559 and used to transform *A. tumefaciens*. Cotelydonary petioles of *B. napus* may be transformed with the recombinant binary vector as described in Moloney et al. (1989, Plant Cell Reports, 8: 238-242).

30 Oil body extracts from transgenic *B. napus* plants may analysed by Western blotting using an anti oleosin antibody for the presence of the fusion protein.

b) **Combination of Two Oleosin Fusion Proteins to Release a Protein Product from Oil Bodies.** Two different oil body protein fusions associated with oil bodies can be used as a means to obtain a final product. For example, a transgenic *B. napus* may be obtained which contains a gene that

5 comprises the GUS enzyme fused to the 3' coding sequence of oleosin separated by a collagenase protease recognition site. Oil bodies may be obtained from the seed of this plant. These oil bodies can be mixed with the oil bodies described above, which contains collagenase fused to oleosin. The collagenase activity of the oleosin/collagenase fusion protein oil bodies can release the GUS enzyme

10 from the oleosin/GUS fusion proteins oil bodies. The GUS enzyme remains in the aqueous phase after removal of the oil bodies. No collagenase enzyme or contaminating oleosins will remain associated with the purified GUS enzyme illustrating the utility of the invention in obtaining easily purified proteins.

c) **Expression of a Oleosin/Phytase fusion protein in *B. napus*.** A

15 microbial phytase from a *Aspergillus* may be isolated based on the published sequence (van Gorcom et al, European Patent Application 90202565.9, publication number 0 420 358 A1). This gene can be fused to the carboxy terminus of the oleosin protein using techniques described above and a collagenase recognition protease cleave site may be included to allow for

20 separation of the phytase from the oil body if desired. The construct may contain, in the following order, the promoter region of the *Arabidopsis* oleosin gene, the coding sequence of the oleosin protein including the intron, a collagenase cleavage site and the phytase gene followed by the nos terminator polyadenylation signal. The construct can be inserted into the binary plasmid

25 Bin 19 and the resultant plasmid introduced into *Agrobacterium*. The resulting strain can be used to transform *B. napus*. The seed of the transgenic plants will contain phytase activity. The phytase activity will be associated with the oil body fraction. The phytase activity is useful for the enhancement of meal for monogastric animal feed. The phytase may be purified by treatment with

30 collagenase as described in a), or the transgenic seed may be used as a feed additive.

d) **Expression of a Oleosin/Glucose isomerase.** The enzyme glucose isomerase can be expressed as a oleosin fusion protein by joining the coding sequence for the enzyme, (for example, described by Wilhelm Hollenberg, 1985, Nucl. Acid. Res. 13:5717-5722) to the oleosin protein as described above. The 5 construct may be used to transform *B. napus*.

e) **Expression of a Oleosin/High Lysine Fusion Protein.** In order to increase the lysine content of transgenic seeds, a polylysine oligonucleotide may be added to the 3' coding region of the oleosin gene. For example, a repetitive oligonucleotide encoding a polylysine coding sequence can be made by 10 synthesizing a (AAG)<sub>20</sub> oligonucleotide that is joined to the 3' coding region of the oleosin gene by replacement of the hirudin coding sequence contained within pCBOGHIRT plasmid described above in example8 with the polylysine oligonucleotide through the use of cohesive restriction termini. The construct may contain, in the following order, the promoter region of the *Arabidopsis* 15 oleosin gene, the coding sequence of the oleosin protein including the intron, 20 codons for the amino acid lysine followed by the nos terminator polyadenylation signal. The construct may be inserted into the binary plasmid Bin 19 and the resultant plasmid may be introduced into the *Agrobacterium*. The resulting strain can be used to transform *B.napus*.

f) **Expression of a Fungicidal Protein as an Oleosin Fusion Protein.** 20 As a further example of the invention, an oleosin fusion protein may be constructed which encodes a protein that is toxic to fungi. For example, the gene for the enzyme chitinase isolated from tobacco (Melchers et al, 1994, Plant Journal 5:469-480) may be fused to the 3' coding region of oleosin under the 25 control of the native oleosin promoter. Included in this construct may be an oligonucleotide that encodes a collagenase recognition site located between the oleosin and chitinase coding regions. The expression of this construct will result in the production of a oleosin/chitinase fusion protein from which the chitinase enzyme can be released from the oleosin by treatment with collagenase. To this 30 construct may be added a second chimeric gene capable of expression of a collagenase enzyme during seed germination. This second gene can comprise approximately 1.5 Kb of the 5' promoter region for isocitrate lyase, the

collagenase coding sequence of *Vibrio alginolyticus* (Takeuchi et al., 1992, Biochemical Journal, 281:703-708) and the nos terminator. Isocitrate lyase is a glyoxysomal enzyme expressed under transcriptional control during early stages of seed germination (Comai et al., 1989, The Plant Cell, 1:293-300). This 5 second construct therefore will express collagenase during the germination of the seed and mobilization of the oil body reserves. Expression of isocitrate lyase is restricted to germination and is not expressed in developing seeds. This second gene, joined to the oleosin/chitinase gene can be inserted into the binary vector Bin 19. The resultant vector may be introduced into *Agrobacterium* and 10 used to transform *Brassica napus* plants. It is noted that the two genes may also be introduced independently or in two different plants which are then combined through sexual crossing. Seed from transgenic plants would be collected and tested for resistance to fungi.

g) **Expression of an Oleosin Fusion Protein that Provides Protection 15 from Insect Predation.** As a further example of the invention, a fusion oleosin protein may be constructed which encodes a protein toxic to foraging insects. For example, the gene for cowpea trypsin inhibitor (Hilder et al., 1987, Nature, 330:160-163) may be used to replace the chitinase gene described in e). The expression of this construct will result in the production of a oleosin/trypsin 20 inhibitor fusion protein from which the trypsin inhibitor can be released from the oleosin by treatment with collagenase. By replacement of the chitinase gene in e) with the trypsin inhibitor, the construct also contains the collagenase gene under control of the germination specific promoter from the isocitrate lyase 25 gene. This construct may be inserted into the binary vector Bin19. The resultant vector can be introduced into *Agrobacterium* and used to transform *Brassica napus* plants. Seed from transgenic plants were collected and tested for resistance to insect predation.

h) **Expression of an Enzyme to Alter Secondary Metabolites in 30 Seeds.** In order to alter specific secondary metabolites in the seed, an enzyme encoding tryptophan decarboxylase (TDC) can be expressed in the seed as a fusion to oleosin. This particular enzyme (DeLuca et al., 1989, Proc. Natl. Acad. Sci. USA, 86:2582-2586), redirects tryptophan into tryptamine and causes a

depletion of tryptophan derived glucosinolates. This lowers the amount of the antinutritional glucosinolates in the seed and provides a means to further reduce glucosinolate production in crucifer plant species. To accomplish this, a fusion protein may be constructed between the TDC gene and the oleosin coding 5 region. The construct may contain, in the following order, the promoter region of the *Arabidopsis* oleosin gene, the coding sequence of the oleosin protein including the intron, the TDC gene followed by the nos terminator polyadenylation signal. The construct may be inserted into the binary plasmid Bin 19 and the resultant plasmid introduced into *Agrobacterium*. The resulting 10 strain can be used to transform *B.napus*.

i) **Expression of Heterologous Proteins in Mammalian Cells.** The oil body protein - heterologous protein fusion may also be prepared in mammalian host cells. For example, an oleosin/GUS fusion may be inserted 15 into a mammalian expression vector and introduced into mammalian cells as described below.

Expression of an oleosin/GUS fusion in mammalian cells would require the cloning of the GUS gene as described in example 17 in commercially available mammalian expression vectors. For example, mammalian expression vectors pMSG, pSVL SV40, pCH 110, (all available from Pharmacia code No. 27- 20 4506-01, 27-4509-01 and 27-4508-1 respectively) may be used. The oleosin/GUS fusion gene may be fused in the plasmid. These plasmids can be introduced into mammalian cells using established protocols (See eg. Introduction of DNA into mammalian cells (1995) Current Protocols in Molecular Biology, Ausubel et al. (ed) Supplement 29, Section 9). Accumulation of the oleosin/GUS transcript in 25 mammalian cells can be determined after preparation of mammalian cell RNA (See eg. Direct analysis of RNA after transfection (1995) Current Protocols in Molecular Biology, Ausubel et al. (ed) Supplement 29, Section 9.8), northern blotting, and hybridization of this northern blot to a <sup>32</sup>P labelled *Brassica* oleosin cDNA as described in Example 18. After preparation of a total protein extract 30 from the transfected mammalian cell culture, GUS activity can be measured, demonstrating the accumulation of the oleosin/GUS protein. Alternatively,

immunoblotting can be performed on this protein extract using commercially available GUS antibodies and/or oleosin antibodies.

**Table I. Expression of *Arabidopsis oleosin* chimeric promoter constructs in transgenic *Brassica napus*.**

| Promoter Construct (GUS fusion) | Expression of GUS Activity (pmol/MU/mg protein/min) |      |      |      |                       |
|---------------------------------|---|------|------|------|-----------------------|
|                                 | Early Seed (torpedo)                                | Root | Leaf | Stem | Late Seed (cotyledon) |
| 2500                            | 7709  | 444  | 47   | 88   | 11607                 |
| 1200                            | 1795  | -    | -    | -    | 8980                  |
| 800                             | 475   | -    | -    | -    | 7130                  |
| 600                             | 144   | -    | -    | -    | 1365                  |
| 200                             | 65  | 260  | 6    | 26   | 11                    |
| control                         | 14  | 300  | 6    | 30   | 14                    |

5 Oleosin promoter-GUS fusions were constructed as described in example 3. Included are GUS values obtained from a control non-transformed plant. A (-) indicated the tissue was not tested. Units are picomoles of methyl umbelliferone (product) per mg protein per minute.

10 **Table II. Expression of *Arabidopsis oleosin* chimeric promoter constructs in transgenic tobacco (*Nicotiana tabacum*).**

| Promoter Constructs (GUS fusions) | GUS Activity in Seeds (pmol/MU/mg protein/min) |
|-----------------------------------|--|
| 2500                              | 11330  |
| 800                               | 10970  |
| Control                           | 0  |

Oleosin promoter-GUS fusions were constructed as described in example 3. Included are GUS values obtained from a control non-transformed plant. Units 15 are picomoles of methyl umbelliferone (product) per mg protein per minute.

**Table III. Specific partitioning of GUS/oleosin fusions into oil bodies when expressed in transgenic *Brassica napus* plants.**

| Plant Number | Percent GUS Activity in Oil Bodies (%) | GUS Activity in Oil bodies | GUS Activity 100,000 X g Supernatant | GUS Activity in 100,000 X g Pellet |
|--------------|--|----------------------------|--------------------------------------|------------------------------------|
| A1           | 88                                     | 493                        | 1                                    | 67                                 |
| B7           | 90                                     | 243                        | 5                                    | 22                                 |
| control      | 0                                      | 0                          | 0                                    | 0                                  |

Plants were transformed with an oleosin/GUS fusion protein under the control of the *Arabidopsis* oleosin promoter. Transformed seeds were obtained and fractionated. The initial fractionation consisted of grinding the seeds in 1.5 mL of buffer A consisting of 15 mM Tricine-KOH, pH 7.5, 10 mM KCl, 1 mM Mg Cl<sub>2</sub>, 1 5 mM EDTA, 100 mM sucrose followed by centrifugation at 14,000 X g for 15 minutes at 4°C. From this three fractions were obtained consisting of a floating oil body layer, an aqueous layer and a pellet. The oil body fraction was recovered and assayed for GUS activity. The remaining aqueous phase was further centrifuged for 2 hours at 100,000 X g. The pellet and supernatant from 10 this centrifugation was also tested for GUS activity. Units are nmol MU per mg protein per min.

15 **Table IV. Cleavage of GUS enzyme from oleosin/GUS fusions associated with oil bodies derived from transgenic *Brassica napus* containing an oleosin/GUS fusion protein.**

| GUS Activity<br>(nmol product/mg protein/min) |                 |                |            |
|---|-----------------|----------------|------------|
| Fraction                                      | Before Cleavage | After Cleavage | % Activity |
| Oil bodies                                    | 113             | 26.4           | 24         |
| 100,000 X g supernatant                       | 14.3            | 83.6           | 76         |
| 100,000 X g pellet                            | 15.7            | -              | -          |

Oil bodies containing an oleosin/GUS fusion protein were subjected to cleavage using the endopeptidase thrombin as described in example 5. Values shown are 20 GUS activities before and after cleavage with thrombin. The values are also expressed as a percentage of total GUS activity released following enzyme fusion. Units are nmol methyl umbelliflorone per mg protein/min.

25 **Table V. Reuse of oil body associated enzymatic activities.**

| # Times Oil Bodies Washed | % GUS Activity |             |
|---------------------------|----------------|-------------|
|                           | Oil bodies     | Supernatant |
| 1                         | 100            | 8 ± 5       |
| 2                         | 118 ± 7        | 5 ± 3       |
| 3                         | 115 ± 8        | 3 ± 4       |
| 4                         | 119 ± 8        | 1 ± 20      |

Oil bodies containing an oleosin/GUS protein were isolated from the seeds of transgenic *Brassica napus*. The oil bodies were added to the fluorometric GUS substrate MUG and allowed to react for one hour. The oil bodies were then recovered and added to a new tube containing the substrate and allowed to react for one hour again. This process was repeated a total of four times. The table illustrates the reusable activity of the GUS enzyme while still associated with the oil bodies. Values are normalized to 100% as the GUS activity of original oil body isolates.

10 **Table VI. Recovery of active hirudin following synthesis of hirudin in plant seeds.**

| Treatment                  | Thrombin Units Per Assay | Antithrombin Units per mg Oil Body Proteins |
|----------------------------|--------------------------|---|
| Buffer only                | 0.143                    | 0   |
| Wild-type seed             | 0.146                    | 0   |
| Wild-type seed + factor Xa | 0.140                    | <0.001                                      |
| Transformed (uncut)        | 0.140                    | <0.001                                      |
| Transformed + factor Xa    | 0.0065                   | 0.55  |

15 Oil bodies containing a hirudin/GUS fusion protein were isolated according to the method and treated with the endoprotease Factor Xa inhibition assay using N-p tosyl-gly-pro-arg-p-nitro anilide (Sigma). Hirudin activity was measured by the use of a thrombin in the method of Dodt et al (1984, FEBS Lett. 65, 180-183). Hirudin activity is expressed as thrombin units per assay in presence of 255  $\mu$ g of oil body proteins, and also as antithrombin units per mg oil body protein.

20 **Table VII: Expression of active oleosin/GUS fusions in *E. coli*.**

| Plasmid    | Gus Activity |
|------------|--------------|
| pKK233-2   | 2.5          |
| pKKoleoGUS | 3.1          |
| pKKoleoGUS | 28.1         |
| pkkGUS     | 118.2        |

As described in example 22, oleosin/GUS fusions were expressed in *E. coli*. Cells were grown, induced with ITPG and GUS activity measured.

**Table VIII: GUS activity of total extracts of untransformed *S. Cerevisiae* strain 5 1788, transformed with M1830 and M1830oleoGUS**

| <b><i>S. Cerevisiae</i> strain 1788</b> | <b>Specific Activity (pmol MU.min<sup>-1</sup>.μg prot<sup>-1</sup>)</b> |
|---|--|
| untransformed                           | 0.001  |
| transformed with M1830                  | 0.001  |
| Transformed with M1830 OleosinGUS       | 41.3   |

While the present invention has been described with reference 10 to what are presently considered to be the preferred examples, it is to be understood that the invention is not limited to the disclosed examples. To the contrary, the invention is intended to cover various modifications and equivalent arrangements included within the spirit and scope of the appended claims.

15 All publications, patents and patent applications are herein incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.